



HAL
open science

Comprehensive Sample Preparation Device for Multi-Omics Analysis

Ranjith Kumar Ravi Kumar

► **To cite this version:**

Ranjith Kumar Ravi Kumar. Comprehensive Sample Preparation Device for Multi-Omics Analysis. Analytical chemistry. Université Paris sciences et lettres, 2023. English. NNT : 2023UPSLS005 . tel-04081248

HAL Id: tel-04081248

<https://pastel.hal.science/tel-04081248>

Submitted on 25 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à Ecole Supérieure de Physique et de Chimie Industrielles
de la Ville de Paris : UMR 8249 – Plasticité du Cerveau

**Dispositif Complet de Préparation d'Echantillons pour
l'Analyse Multi-Omique**

Comprehensive Sample Preparation Device for Multi-Omics
Analysis

Soutenue par

Ranjith Kumar RAVI KUMAR

Le 08 mars 2023

Ecole doctorale n° 388

**Chimie physique et chimie
analytique de Paris centre**

Spécialité

Chimie analytique

Composition du jury :

Andrew, GRIFFITH Pr, ESPCI-PSL Paris	<i>Président du Jury</i>
Martial, MARBOUTY CR, Institut Pasteur	<i>Rapporteur</i>
Virginie, REDEKER CR, Université Paris-Saclay	<i>Rapporteur</i>
Pascal, COSETTE Pr, Université Rouen	<i>Examineur</i>
Joelle, VINH DR, ESPCI-PSL Paris	<i>Examineur</i>
Yann, VERDIER MC, ESPCI-PSL Paris	<i>Directeur de thèse</i>

தாமின் புறுவது உலகின் புறக்கண்டு

காழுறுவர் கற்றறிந் தார்.

*When the learned see that the learning that delights them delights the world
as well, they love learning even more.*

-Couplet 399, Thirukural

Research is to see what everybody else has seen,

and to think what nobody else has thought.

-Albert Szent-Gyorgyi

Acknowledgments

Members of the thesis monitoring committee and jury of the thesis along with the reporters are acknowledged for their valuable comments about the project.

I wish to thank and share the successful completion of my thesis with all members of the SMBP laboratory. Joëlle Vinh and Giovanni Chiappetta for contributing to many interesting scientific discussions and for providing a convivial workspace. Yann Verdier for encouraging me through the three years and always having confidence in my research ideas. Massamba Mbacké Ndiaye for training me in the concepts of ChipFilter and challenging me to achieve more. Emmanuelle Demey and Iman Haddad for being my saviours from difficulties with LC-MS and bioinformatics. Alexandra Emmanuel, Nicolas Eskenazi, and Jean Bouvet for supporting experiments. Zeyuan Xu, Wafa Hechiche, and Xhesica Limaj for their friendship. All interns allowed me to train them or discuss my work. I wish to give my thanks to Karan for all the nice discussions we shared in Tamil.

I want to thank Alexandra Marynets, Virginie Cardoso, and members of the Europe department-ESPCI for their support through the program. Ines Maury along with other members of the entrepreneurship department of PSL and members of P2i for the training. Members of YouRPSL (Chandra, Oscar, Nazanin, Jean-Baptiste, and others) for allowing me to improve my skills in managing the activities of the student club.

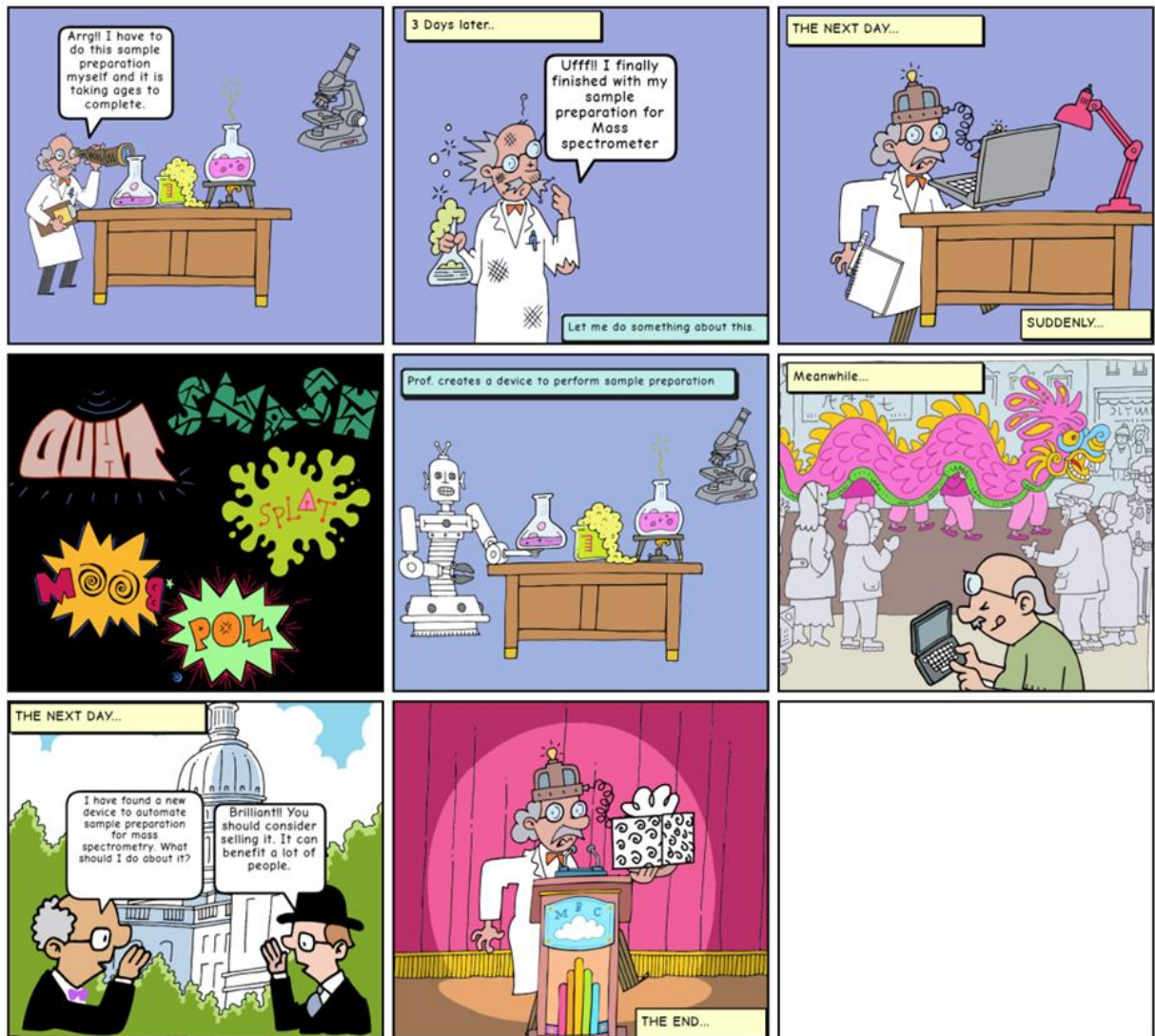
The writing of my thesis and completion of experiments would have not been fun or possible without constant encouragement from Sandy. Thank you for tolerating my tantrums and for the love you give me. Big thanks for the references.

I want to thank my friends from MINA, CIUP- Zayneb, Takwa, Oumayma, Hadhami, Humna, Razane, Patrick, Parth, Pedro, Clement, Jin, Ivan and rest for making my stay in Paris joyful and ensuring my wellbeing. Ian, Sharbat, Gaspard, Nikhil, Anudeep, and Kannan for their friendship.

I wish to give this success to my parents (Kavitha and Ravi) and brother (Rohith). The freedom and faith they had in me allowed me to cross oceans to reach this country and survive these last few years by myself. Thank you very much for your love and prayers. I also wish to thank my friend Arvind and his family (Akila and Balasubramaniam) for their benevolence. I wish to thank Dr. Sailaja Elchuri and her family for encouraging me to come to Paris for my Ph.D. and for all the scientific support she has given me before and during my Ph.D. I thank my friends living outside France – Machris, Muthu, Karthick, Sai, Kirti, Madhumitha, Rashmi, Harish, Senthil, Hemanth, and many others for having stayed during my important phases and always supporting me. Finally, I wish to extend my gratitude to everyone who had an opportunity to influence my studies and life in Europe.

The project is illustrated as a comic,

A scientist tired of preparing samples for proteomic analysis decides to develop an automated device that allows him to prepare the samples in his absence (who wants to miss the festival!!)



Summary in French

Dispositif complet de préparation d'échantillons pour l'analyse multi-omique

Introduction

Depuis la découverte des micro-organismes par Leeuwenhoek (Lane, 2015) au 16ème siècle, chaque époque a été marquée par des avancées scientifiques importantes comme la pasteurisation (Berche, 2012), les antibiotiques (Aminov, 2010), la découverte des archées (Woese & Fox, 1977), les biofilms (Flemming, 1993) et le séquençage du génome bactérien (Fleischmann et al., 1995). Ces découvertes ont été possibles grâce à des développements technologiques tels que la microscopie, les méthodes de culture *in vitro*, la spectrométrie de masse, le séquençage Sanger, la réaction en chaîne par polymérase et le séquençage de nouvelle génération. L'émergence de nouvelles technologies et les découvertes scientifiques subséquentes ont conduit à l'élargissement de la définition du microbiome qui a été initialement proposée par Whipps et al. (Whipps et al., 1988) et qui es la suivante :

"Un cadre écologique commode pour examiner les systèmes de biocontrôle est celui du microbiome. Celui-ci peut être défini comme une communauté microbienne caractéristique occupant un habitat raisonnablement bien défini et présentant des propriétés physio-chimiques distinctes. Le terme ne se réfère donc pas seulement aux microorganismes impliqués mais englobe également leur théâtre d'activité".

Pour une meilleure compréhension, les différentes études sur le microbiome peuvent être séparées en deux catégories : environnementale (en dehors du corps de l'hôte) et humaine/hôte (à l'intérieur ou sur le corps de l'hôte). Cette classification est basée sur la conséquence probable que le microbiome peut avoir sur l'environnement et la santé de l'homme.

Le microbiome de l'environnement est très diversifié et distinct en fonction de son écosystème. Il se développe et dirige les activités exclusives à leur environnement (comme la formation de nuages de pluie dans l'atmosphère). L'identification du type de micro-organismes, des métabolismes et de leurs interactions est essentielle pour les chercheurs qui étudient ces microbiomes environnementaux. Les principaux sites de l'environnement où les communautés microbiennes ont été étudiées comprennent les eaux marines ou douces, les sols, les eaux usées ou les boues activées, l'atmosphère et les sites d'activité humaine comme les mines.

Les micro-organismes ont prospéré partout sur la terre, y compris chez les humains. Contrairement à l'idée erronée mais répandue selon laquelle ils ne peuvent être

que nuisibles, nous nous épanouissons en fait dans la coopération. Ils participent à différentes fonctions métaboliques comme l'absorption, la digestion et l'élimination des déchets. Dans le corps humain, les communautés microbiennes sont susceptibles de se trouver dans la peau (Grice et al., 2009), la conjonctive (Cui et al., 2013), la cavité buccale (Floyd et al., 2010), le tractus gastro-intestinal (intestin) (Arumugam et al., 2011), les voies urinaires (Perez-Carrasco et al., 2021), le vagin (France et al., 2022), le placenta (Aagaard et al., 2014), l'utérus (Baker et al., 2018), la cavité nasale (Bassis et al., 2014), le poumon (Moffatt et al., 2017) et les voies biliaires (Binda et al., 2022). Les études ont été réalisées en évaluant les biofluides tels que la sueur, les écouillons de peau, les larmes, la salive, les fèces, les mucus, l'urine ou le sang collectés auprès de différents donneurs ou par des biopsies dans des cas spécifiques.

Même si les microbiotes environnementaux et humains/hôtes sont différents, les éléments structurels microbiens (acides nucléiques, protéines et métabolites) sont biochimiquement identiques. Par conséquent, les technologies disponibles pour étudier ces biomolécules sont semblables pour les deux groupes. Toutefois, la différence se fera sentir au stade pré-analytique, lorsque les cellules microbiennes constituant la communauté interagissent (avec les cellules hôtes, les produits chimiques de l'environnement, les facteurs physiques comme le pH ou la salinité, etc). Ces facteurs structurels internes et externes déterminent la complexité de l'étude d'un microbiome spécifique.

Le choix de l'approche est déterminé par les questions auxquelles on souhaite répondre. Les trois grandes questions auxquelles les technologies disponibles permettent actuellement de répondre sont les suivantes : (i) qui est là ? (ii) que peuvent-ils faire ? (iii) que font-ils ? Dans de nombreux cas, il est possible de répondre à plusieurs de ces questions avec une sensibilité limitée ou de répondre en profondeur à une seule question en intégrant plus d'un séquençage d'acides nucléiques, de protéines ou de métabolites.

La métagénomique peut être définie comme l'étude du matériel génétique récupéré directement des communautés présentes dans les échantillons environnementaux ou cliniques. Après l'isolement du matériel génétique (ADN génomique, plasmides ou petits fragments d'ADN) de la matrice ou des cellules, il est généralement séquencé par des méthodes de séquençage *shotgun* à haut débit. Les séquences générées sont utilisées pour construire des génomes assemblés de métagénome (MAGs). Les génomes de communautés microbiennes cultivées, isolées de l'environnement ou d'isolats cliniques, peuvent également être séquencés de manière similaire afin d'identifier les perturbations du système.

La métaprotéomique peut être définie comme la caractérisation à grande échelle de l'ensemble des protéines du microbiote environnemental à un moment donné. Cette définition a été proposée en 2006 (Wilmes & Bond, 2006) et depuis, la recherche sur ce sujet a augmenté de façon exponentielle. Les progrès de la spectrométrie de masse et des techniques de séparation comme la chromatographie liquide y ont contribué de manière significative. La métaprotéomique est étroitement liée à la métagénomique et est souvent utilisée comme une technique complémentaire pour caractériser le microbiome. C'est la raison pour laquelle elle est souvent appelée "protéogénomique".

La multiomique ou omique intégrative ou pan-omique est une stratégie visant à intégrer plus d'une étude omique qui comprend la génomique, la transcriptomique, la protéomique, la métabolomique et la microbiomique. L'intégration de plusieurs ensembles de données pour un même échantillon permet de cartographier l'état de l'organisme à différentes couches qui répondront aux questions relatives à l'origine, la fonction, la régulation et l'interaction. Cette connaissance suprême de l'activité globale de la cellule dans un certain environnement permet de prédire son comportement. Les études sur le microbiome sont les plus grands bénéficiaires de la multiomique. Contrairement aux autres études qui utilisent des organismes modèles pour comprendre la dysbiose, le microbiome traite d'organismes non modèles et de leur fonction dans un certain environnement. L'implication de plus d'une étude omique est nécessaire pour identifier le type microbien et associer leurs fonctions.

Les étapes de la métagénomique ou de la métaprotéomique sont la préparation pré-analytique des échantillons, le séquençage et l'analyse des données. La première étape impliquant la préparation de l'échantillon est spécifique pour chaque famille de biomolécules et est très importante pour déterminer la qualité du séquençage. Plusieurs flux de travail ont été établis pour isoler et préparer les biomolécules nécessaires à la métagénomique ou à la métaprotéomique. Malgré les nombreux flux de travail disponibles, la métagénomique et la métaprotéomique souffrent de plusieurs limitations à l'étape de la préparation de l'échantillon, comme la lyse et l'extraction de cellules complètes spécifiques à l'échantillon, des protocoles efficaces pour le fractionnement du protéome subcellulaire (membrane, fractions extracellulaires et solubles), le traitement d'un faible nombre de cellules, la manipulation d'échantillons pathogènes et l'automatisation. Pour surmonter ces limitations, ce travail propose l'utilisation de la technologie microfluidique.

La technologie microfluidique utilise le contrôle et la manipulation précis de fluides qui sont géométriquement limités à un petit espace (inférieur au millimètre) dans lequel les forces de surface dominent les forces volumétriques. Les systèmes

microfluidiques peuvent transporter, mélanger, séparer et traiter des fluides. Ils trouvent des applications dans des situations qui nécessitent un multiplexage, une automatisation et un criblage à haut débit. Cela rend la microfluidique indispensable pour les applications de chimie analytique. Le système est constitué de composants tels que des micropompes et des microvalves, utiles respectivement pour l'alimentation en fluides et la régulation de la direction du flux. Dans la recherche sur le microbiome, des défis se posent à différentes étapes du traitement des échantillons, notamment la collecte de cellules dans des environnements très dispersés comme les aérosols, la lyse cellulaire, le traitement des échantillons unicellulaires et l'automatisation. La collecte et le traitement des échantillons étant cruciaux pour la préservation de la biomasse et l'analyse ultérieure, plusieurs technologies microfluidiques ont été développées à cet effet. La technologie microfluidique a été utilisée pour développer des puces capables d'effectuer la préconcentration de protéines ou de peptides pour compléter les étapes de préparation préanalytique des échantillons. En génomique, outre la lyse cellulaire, elle a trouvé des applications pour la séparation, la détection et l'amplification des acides nucléiques. Dans le domaine de la préparation et de la préconcentration des échantillons, la technologie microfluidique basée sur les microbilles, les monolithes et les membranes a été largement utilisée. Le dispositif microfluidique utilisé dans ce travail, nommé *ChipFilter* (« puce à filtre ») utilise une membrane neutre avec une coupure de poids moléculaire de 10kDa. Il a été démontré que ce dispositif pouvait effectuer toutes les étapes de préparation des échantillons après la lyse afin de faciliter le couplage direct avec le LC-MS/MS pour l'analyse. Plus de 4000 protéines ont été identifiées à partir de 28 000 cellules eucaryotes HT29 (Ndiaye et al., 2020).

Objectifs de l'étude

Ce travail vise à évaluer et développer l'utilisation de le *ChipFilter* pour l'étude de microbiomes.

Le microbiome intestinal est un modèle largement étudié qui englobe la métagénomique, la métatranscriptomique, la métaprotéomique et la métabolomique. Les bases de données de référence pour 4 644 espèces de l'intestin humain ont été établies (Almeida et al., 2021), ce qui en fait un système modèle idéal pour tester de nouvelles stratégies ou technologies pour le microbiome et les omiques. Le microbiome est généralement isolé à partir des selles de l'hôte, car les spécimens sont collectés naturellement, non invasifs et peuvent être échantillonnés à plusieurs reprises (Tang et al., 2020). Par ailleurs, plusieurs systèmes modèles non humains, y compris des modèles in vitro et des modèles animaux, ont été développés pour étudier les fonctions du microbiome intestinal.

1. Développer le ChipFilter pour le traitement microbien de cellules entières : y compris le chargement et la lyse des cellules dans le dispositif.
2. Évaluer les performances du ChipFilter pour l'analyse des protéines d'une population cellulaire mixte : pour cela, nous comparerons les performances du ChipFilter et des méthodes existantes ;
3. Développer un *workflow* pour collecter les acides nucléiques des cellules lysées dans le ChipFilter, afin d'effectuer des analyses protéomiques et génomiques sur le même échantillon.
4. Valider les deux types d'analyse sur un microbiome standard de l'intestin.
5. Évaluer l'adaptabilité de la puce pour l'étude d'échantillons de microbiome intestinale de souris et de microbiomes environnementaux.

1. Développement du ChipFilter pour le traitement microbien de cellules entières

L'étude actuelle démontrera de nouveaux flux de travail utilisant le dispositif ChipFilter pour effectuer une préparation automatisée et intégrée des échantillons pour la métaprotéomique et l'isolement séquentiel des acides nucléiques. Le ChipFilter a été développé précédemment dans le laboratoire pour effectuer la préparation d'échantillons pour la protéomique bottom-up. Dans le présent travail, j'ai d'abord mis à niveau cette technologie pour effectuer la lyse de cellules bactériennes et fongiques, la purification des protéines, la protéolyse et l'isolement des acides nucléiques. Cinq différents flux de travail démontrés permettent la possibilité d'une préparation efficace des échantillons pour la métaprotéomique seule ou la méta-protéomique collective. Les flux 1-3 ont permis la préparation d'échantillons avec lyse cellulaire, purification des protéines, protéolyse et récupération des peptides pour LC-MS/MS. Le workflow 1 permet l'élution directe des peptides du ChipFilter vers la colonne de piégeage dans un LC, tandis que le workflow 2 permet de piéger les peptides dans un flacon. Le flux de travail 3 permet d'avoir une durée de protéolyse plus courte. Le tampon de lyse modifié dans le flux de travail 5 a permis d'augmenter la lyse cellulaire des bactéries Gram +ve. Le flux de travail 4 permet de récupérer les acides nucléiques après la récupération des peptides. La démonstration des flux de travail a été étendue pour être appliquée à la préparation d'échantillons microbiens.

Le flux de travail 1 du ChipFilter a démontré sa capacité à traiter des cellules bactériennes et fongiques pour identifier les peptides et les protéines correspondantes. Le nombre de cellules était d'environ un million pour chaque type de cellule. Le nombre de protéines identifiées pour *Escherichia coli*, *Bacillus subtilis* et *Saccharomyces cerevisiae*

est de $1\ 121 \pm 58$, 824 ± 50 et $1\ 339 \pm 44$ respectivement ; tandis que le nombre de peptides était de $4\ 694 \pm 340$, $5\ 802 \pm 463$ et $7\ 959 \pm 412$ respectivement. Le nombre de PSM identifiés est respectivement de $39\ 466 \pm 1\ 757$, $36\ 302 \pm 1\ 083$, et $75\ 096 \pm 1\ 120$. Les identifications des peptides spécifiques à l'espèce prouvent que les cellules bactériennes et fongiques peuvent être traitées sans prétraitement par le procédé ChipFilter.

La procédure de fabrication du ChipFilter est expliquée.

Différents flux de travail basés sur le ChipFilter pour le traitement des échantillons pour la protéomique et l'isolation des acides nucléiques sont été décrits en détail. Des développements ont été réalisés pour surmonter les contaminants qui sont causés par le chargement de cellules entières avec le système Fluigent.

La lyse des cellules microbiennes avec un tampon de lyse à base de détergent non ionique dans le système ChipFilter a été réalisée avec succès.

La préparation d'échantillons pour *Escherichia coli*, *Bacillus subtilis* et *Saccharomyces cerevisiae* à partir de cellules entières a été réalisée avec succès et analysée par LC-MS/MS.

2. Évaluation des performances de ChipFilter

Ensuite, il s'agit d'évaluer les performances de la différence entre le flux de travail ChipFilter et les flux de travail existants comme la préparation d'échantillons assistée par filtre (FASP), in-gel et in-solution. Je compare les flux de travail pour la préparation d'échantillons de cellules mélangées en termes de temps pris pour chaque préparation, l'identification des protéines ou des peptides, la distribution des protéines entre les trois espèces, l'efficacité de la manipulation avec des échantillons à faible nombre de cellules et les propriétés des peptides, y compris les clivages manqués, l'hydrophobie, la longueur, la masse théorique des ions positifs et les PSM générés. Enfin, la répétabilité du flux de travail ChipFilter a été comparée à d'autres flux de travail et entre différents nombres de cellules pour les cellules mixtes. Cette section est très importante pour établir la base fondamentale de la nouveauté et de l'amélioration de l'identification des protéines par le flux de préparation des échantillons ChipFilter.

L'identification des protéines et des peptides est très importante pour évaluer la performance des flux de travail préanalytiques. Pour le flux de travail ChipFilter, une identification moyenne des protéines de $1\ 999 \pm 54$ a été atteinte alors que les autres flux de travail ont atteint $1\ 818 \pm 91$ (mFASP), $1\ 504 \pm 22$ (in-gel) et $1\ 665 \pm 15$ (in-solution) respectivement lorsque trois millions de cellules ont été prises. Le flux de travail ChipFilter

a une identification significativement plus élevée que le flux de travail in-gel ou in-solution et est aussi compétent que mFASP pour environ trois millions de cellules. Les identifications de peptides sont de $9\,770 \pm 108$ (ChipFilter), $8\,197 \pm 435$ (mFASP), $8\,218 \pm 262$ (in-gel) et $7\,805 \pm 178$ (in-solution). L'identification est nettement plus élevée pour le flux de travail ChipFilter que pour les autres flux de travail. L'efficacité de la catalyse dans un espace restreint et la réduction des pertes de matériaux peuvent être attribuées au succès de l'identification plus élevée pour le flux de travail ChipFilter. En outre, la distribution des protéines et des peptides entre les flux de travail a été étudiée. En ce qui concerne la distribution des protéines, on constate que 17,5% (598) du total des protéines ont été identifiées exclusivement par le flux de travail ChipFilter, tandis que les autres flux de travail avaient moins de 5,1% de protéines identifiées exclusivement par ce flux de travail. Pour la distribution des peptides, 32% (8 395) du total des peptides ont été identifiés uniquement par le flux de travail ChipFilter. Il est intéressant de noter que l'identification peptidique totale commune à tous les flux de travail n'était que de 8,8% (2 309), ce qui indique que les différents flux de travail génèrent de manière caractéristique des fractions peptidiques différentes. La comparaison avec les trois cellules microbiennes indique que le flux de travail ChipFilter identifie le pourcentage le plus élevé pour les bactéries gram-positives *B. subtilis* (12%) et les bactéries gram-négatives *E. coli* (37%) par rapport aux autres flux de travail, alors que les champignons *S. cerevisiae* (51%) ont été enregistrés.

La robustesse du flux de travail ChipFilter a été testée pour les échantillons à faible nombre de cellules. Les résultats montrent que le nombre moyen de protéines et de peptides identifiés est de 163 ± 18 et $1\,162 \pm 126$ respectivement. De plus, l'identification des protéines est plus faible pour le flux de travail ChipFilter que pour le flux de travail in-solution (172 ± 28), mais les identifications de peptides sont plus élevées que pour tous les autres flux de travail. Ensuite, la distribution des protéines indique que le pourcentage de protéines exclusives identifiées par le flux de travail ChipFilter (20,6% (79)) et le flux de travail in-solution (17,5% (67)) est très proche du nombre élevé de cellules. De plus, le pourcentage de protéines communes entre les deux flux de travail est d'environ 45,9%, ce qui suggère que l'identification des protéines entre les deux flux de travail est très similaire. Sur le total des protéines identifiées, seulement 27,9% des protéines n'ont pas été identifiées par le flux de travail ChipFilter. Pour l'identification des peptides, le pourcentage de peptides exclusifs identifiés par le flux de travail ChipFilter est de 45,1% (1 270). C'est trois fois plus que le flux de travail en solution (13,4% (378)). La proportion de peptides non identifiés par le flux de travail ChipFilter est de 30,4% (858). Comme pour le nombre plus élevé de cellules, l'annotation des protéines identifiées appartient aux trois espèces pour le flux de travail ChipFilter. Dans le cas de l'identification de

B. subtilis, le ChipFilter identifie 9% des protéines totales, alors qu'il est inférieur à 4% pour les autres flux de travail. En outre, l'identification des protéines de *S. cerevisiae* pour le flux de travail en solution est de 87% du total des protéines identifiées, ce qui est plus élevé que les autres flux de travail qui ont environ 50% à 60% de cette espèce. Le flux de travail ChipFilter est clairement un flux de travail universel de préparation d'échantillons pour l'identification de protéines de plusieurs espèces en même temps, même avec un faible nombre de cellules.

L'analyse des propriétés physico-chimiques des peptides générés par les différents flux de travail indique que la méthode ChipFilter présente un pourcentage plus élevé de clivages manqués, ce qui donne des peptides avec des acides aminés plus longs et des valeurs MH+ plus élevées que les autres flux de travail. En termes de PSMs générés, le flux de travail ChipFilter avait une plus grande proportion de peptides avec moins de 10 PSMs alors que les autres flux de travail avaient plus de peptides avec plus de 10 PSMs. En ce qui concerne l'indice d'hydrophobie, aucune différence n'a été observée entre les différents flux de travail.

Enfin, l'estimation de la répétabilité entre les différentes répliques en utilisant la méthode LFQ a montré une bonne répétabilité pour le flux de travail ChipFilter mais elle était inférieure à celle des autres flux de travail.

3. Développement d'un workflow pour collecter les acides nucléiques des cellules lysées dans le ChipFilter, afin d'effectuer des analyses protéomiques et génomiques sur les mêmes échantillons.

Les flux de préparation des échantillons exigent plus que l'isolement d'une seule biomolécule, car le principal goulot d'étranglement de la multiomique est la limitation des flux de préparation des échantillons disponibles. De plus, comme nous l'avons vu précédemment, les études sur le microbiome nécessitent l'intégration de plus d'une étude omique. À cet égard, un effort est fait pour comprendre l'isolement séquentiel des acides nucléiques après la digestion des protéines dans le flux de travail ChipFilter. De plus, pour étendre l'utilisation du flux de travail ChipFilter, je décrirai dans ce chapitre l'utilisation de ChipFilter pour isoler les acides nucléiques à partir du même échantillon que celui utilisé pour la préparation des échantillons protéomiques. Cette nouvelle approche pour isoler les protéines et les acides nucléiques d'un même échantillon permet d'effectuer des études protéogénomiques très précises. Dans un premier temps, le workflow 4 pour l'isolation de l'acide nucléique après la protéolyse est expliqué en détail. Ensuite, la validation de l'acide nucléique isolé est faite par électrophorèse sur gel d'agarose pour les mélanges de cellules microbiennes et humaines. Enfin, l'amplification

des acides nucléiques isolés (ADN plasmidique) par réaction en chaîne par polymérase est réalisée.

Il est impératif de connaître les différentes étapes auxquelles les acides nucléiques sont retirés du système ChipFilter. Le nombre de cellules utilisées pour les études protéomiques par le flux de travail ChipFilter peut aller de quelques centaines à quelques millions, mais la quantité d'acides nucléiques obtenue n'est pas entièrement proportionnelle à la densité cellulaire initiale, car on observe une perte d'acides nucléiques au cours des différentes étapes du flux de travail protéomique. On constate que les acides nucléiques sortent du ChipFilter pendant le chargement des cellules, la lyse des cellules et pendant l'élution avec une direction de flux inversée.

Il est nécessaire de connaître la quantité de matériel de départ en termes de nombre de cellules pour estimer les acides nucléiques et les protéines qui peuvent être simultanément utiles pour NGS et MS respectivement. Comme on l'a vu, quelques centaines de cellules microbiennes ont suffi pour identifier plus de 500 protéines, mais ce n'est peut-être pas le cas pour les acides nucléiques en raison de la perte observée précédemment. De plus, l'analyse ultérieure ou l'approche de l'identification de l'acide nucléique déterminera les besoins. Pour être visualisées par électrophorèse sur gel d'agarose, un minimum de 10^6 cellules d'origine procaryote ou eucaryote doit être prélevé. Ceci est nécessaire pour surmonter la perte qui se produit pendant le chargement et la lyse des cellules.

On remarque que l'acide nucléique sort du ChipFilter pendant le chargement des cellules et l'étape de lyse. De plus, la proportion du nombre de cellules injectées ne contribue pas exactement à l'acide nucléique obtenu et il y a une perte. Il est donc nécessaire de comprendre le mécanisme de piégeage de l'acide nucléique à l'intérieur du ChipFilter. On suppose que la rétention des acides nucléiques dans le ChipFilter est indirectement médiée par la rétention des protéines liées aux acides nucléiques telles que les protéines de liaison à l'ADN, l'ARN polymérase et les histones.

L'élution de l'acide nucléique est positionnée après la protéolyse qui impliquera une pléthore de réactions avec des produits chimiques comme les détergents non ioniques, le DTT, l'Iodoacetamide et les enzymes qui peuvent également affecter l'acide nucléique présent dans le système. Il est donc nécessaire de s'assurer qu'ils ont été bien protégés de ces produits chimiques. Les acides nucléiques (ADN et ARN) se sont révélés stables après incubation avec le tampon de lyse I, 20 mM de DTT, 50 mM d'IAA, 50 mM de tampon ABC et 2 μ g de trypsine.

Pour l'étude, trois échantillons différents obtenus à partir de cellules humaines, un mélange de trois micro-organismes modèles et un standard de microbiome intestinal comprenant vingt et une espèces ont été utilisés. Les acides nucléiques de tous les mélanges utilisés ont été obtenus par ChipFilter workflow-4.

4. Validation des deux types d'analyse sur un microbiome standard de l'intestin

Afin de valider fonctionnellement l'intégrité de la fraction d'acide nucléique isolée par le ChipFilter workflow-4 et de démontrer la possibilité d'une analyse en aval par NGS ou PCR, j'ai choisi d'amplifier une région spécifique du plasmide qui était présente dans la souche d'E. coli utilisée dans l'étude. À cette fin, la région codant pour la protéine fluorescente verte du plasmide pGEX-4T-3 exprimé dans E. coli a été ciblée et amplifiée.

Afin d'augmenter la portée de l'identification et d'introduire de la complexité en termes de nombre d'espèces et de nombre de cellules des espèces participantes, un nouveau mélange défini de cellules communément trouvées dans l'intestin humain a été pris. Le second mélange de cellules utilisé dans l'étude n'est pas uniforme en termes de nombre de cellules et comporte davantage d'espèces appartenant aux bactéries (bactéries Gram positives et négatives à la coloration de Gram), aux champignons et aux archées. Une moyenne de $2\,521 \pm 375$ protéines et $6\,099 \pm 1140$ peptides et $18\,339 \pm 1535$ PSMs. Parmi les 21 espèces prises, au moins deux protéines ont été identifiées dans plus de 20 espèces. Même les espèces peu abondantes du mélange ont été identifiées. Seules les cellules archéales, présentes à moins de 0,2 %, n'ont pas été identifiées. Même les bactéries gram +ve, qui sont généralement difficiles à lyser et à obtenir des protéines, ont été identifiées efficacement. Dans le cas d'une bactérie positive à la coloration de Gram, *Bifidobacterium adolescentis*, l'utilisation de lysozyme dans le tampon de lyse, comme décrit dans le workflow 5 de ChipFilter, s'est avérée efficace. En outre, l'estimation de la biomasse basée sur les protéines s'est avérée efficace avec plusieurs espèces, qui correspondaient exactement à la quantité théorique décrite par le fabricant. Enfin, l'acide nucléique a été isolé du même échantillon après élution des peptides et résolu par électrophorèse sur gel d'agarose pour en vérifier la qualité, puis soumis à un séquençage de nouvelle génération.

5. Adaptabilité de la puce pour l'étude d'échantillons de microbiome intestinal de souris et environnemental

Ensuite, la validation du flux de travail a été effectuée sur un échantillon réel du microbiome de l'intestin obtenu à partir de fèces de souris. Pour cette étude, les fèces de souris ont été collectées sur des souris hébergées et traitées comme indiqué précédemment pour obtenir le microbiome. Ensuite, ces cellules ont été lysées par un

prétraitement de battage de billes et un petit volume (un quart) des cellules obtenues a été chargé dans le dispositif et soumis à la préparation de l'échantillon. Afin de générer la base de données pour l'identification des protéines, deux approches ont été développées et ont conduit à l'identification de plus de 650 protéines provenant de plusieurs espèces bactériennes précédemment identifiées dans l'intestin de la souris. Cette étude a prouvé que la méthode utilisant la ChipFilter peut être utilisée pour l'étude d'échantillons du microbiome réels.

Enfin, diverses utilisations du ChipFilter, y compris la collecte d'échantillons à partir d'aérosols, ont été mises en oeuvre. Bien que peu d'expériences aient été réalisées à cet effet, les résultats incluant la modification du ChipFilter pour incorporer des parties auxiliaires afin d'améliorer la collecte d'échantillons dans les bioréacteurs, l'utilisation de pompes externes pour faire circuler les aérosols des bio-composts et de la pièce ambiante ont été faits. Ce travail permettra à l'avenir d'inclure la collecte d'échantillons à partir de sources liquides et aérosols pour un traitement ultérieur des échantillons.

Conclusion

En conclusion, un flux de travail complet de préparation d'échantillons utilisant le ChipFilter microfluidique a été réalisé pour la méta-protéogénomique. La possibilité de ce flux de travail d'automatiser la préparation des échantillons avec le potentiel d'être facilement intégré pour la collecte d'échantillons, d'effectuer des études multi omiques sur le même échantillon simultanément et d'identifier les peptides élevés pour les études de microbiome peut être très utile pour effectuer des études de microbiome à haut débit avec facilité.

Scientific Communications

Seminars:

- A delivered oral presentation titled “Comprehensive sample preparation microfluidic device for metaproteomics” in Analytics 2022. Nantes, France. Sept 2022
- A delivered oral presentation titled “Monitoring microbial communities to conserve ecosystem” during ESPCI-PSL Sustainability days. Paris, France. June 2022
- A delivered oral presentation titled “Microfluidic chip for meta omics” during IPGG CARNOT day. Paris, France. May 2022
- A delivered oral presentation titled “Comprehensive sample preparation microfluidic device for metaproteomics” at the International Metaproteomics symposium. Luxembourg. Sept 2021
- Delivered poster presentation titled “Microfluidics based sample preparation of microorganisms for total proteome analysis by mass spectrometry” in SFSM congress France. June 2021

Grants:

- Funded 11 k€ towards developing the ChipFilter project for metaproteomics under the PSL Pépite Proof of Concept (POC) 2020 projects funded by PSL university and Délégations Régionales à la Recherche et à la Technologie (D.R.R.T.).
- Funded 15 k€ towards developing the ChipFilter project for meta-omics under the Construisons Avec la Recherche Notre Offre Technologique (CARNOT) project call in 2021.

Publication:

- Ravi Kumar, R. K., Ndiaye, M. M., Haddad, I., Vinh, J., & Verdier, Y. (2023). ChipFilter: Microfluidic Based Comprehensive Sample Preparation Methodology for Metaproteomics. *BioRxiv*, 2023.01.18.524548.

Social Media:

- YouTube → My Ph.D. in one minute, “<https://youtu.be/FaJ9ael12E8>”

List of Contents

Topic number	Title	Page number
	Acknowledgments.....	v
	Comic description.....	vii
	Summary in French.....	ix
	Scientific communications.....	xxi
	List of Figures.....	xxvi
	List of Tables.....	xxx
	List of abbreviations.....	xxxii
1.	General Introduction.....	1
1.1.	Microbiome- History and definitions.....	2
1.1.1.	Environmental microbiome.....	4
1.1.2.	Human/Host microbiome.....	7
1.1.3.	Approaches in microbiome studies.....	10
1.2.	Metagenomics.....	13
1.2.1.	Steps involved in nucleic acid sequencing.....	14
1.2.2.	Accomplishments.....	20
1.2.3.	Challenges.....	21
1.3.	Metaproteomics.....	24
1.3.1.	Shotgun proteomics.....	25
1.3.2.	Accomplishments.....	33
1.3.3.	Challenges.....	34
1.4.	Microfluidics technology.....	36
1.4.1.	Different fabrication processes and microfluidic systems.....	36
1.4.2.	Microfluidics for microorganisms.....	40
1.4.3.	Microfluidics for sample collection and processing.....	43
1.4.4.	Microfluidics for genomics and proteomics.....	46
1.5.	Multi-omics approaches in microbiome studies.....	51
1.6.	Understanding the gut microbiome.....	54
1.7.	Scope of the thesis.....	57
2.	Development of ChipFilter workflows for whole-cell microbial sample processing.....	59
2.1.	Introduction.....	60
2.2.	Material and methods.....	61
2.2.1.	Biological material.....	61
2.2.2.	ChipFilter development.....	61
2.2.3.	Workflows for sample processing.....	64
2.2.4.	LC-MS/MS analysis and protein identification.....	68
2.3.	Loading and lysis of microbial cells into the device.....	68

2.4.	Single microbial sample preparation and identification.....	70
2.5.	Peptide elution: Trapping column (WF1) versus vial (WF2).....	72
2.6.	Conclusions.....	75
3.	ChipFilter sample preparation performance assessment with existing workflows.....	76
3.1.	Introduction.....	77
3.2.	Material and methods.....	78
3.2.1.	Biological material and chemicals.....	78
3.2.2.	Methods.....	78
3.2.3.	LC-MS/MS analysis and protein identification.....	81
3.2.4.	Data processing.....	82
3.3.	Experimental time.....	82
3.4.	Protein identification and distribution.....	83
3.5.	Low cell number processing.....	87
3.6.	Peptide level variations.....	89
3.7.	Repeatability.....	93
3.8.	Conclusions.....	95
4.	Proteomic analysis of murine gut microbiome: a proof of concept	96
4.1.	Introduction.....	97
4.2.	Material and methods.....	98
4.2.1.	Biological material.....	98
4.2.2.	Protein extraction.....	98
4.2.3.	LC-MS/MS analysis.....	99
4.2.4.	Data analysis.....	99
4.3.	Choice of the database.....	100
4.4.	Taxonomic analysis.....	102
4.5.	Functional analysis.....	105
4.6.	Conclusions.....	108
5.	Going beyond proteomics.....	109
5.1.	Introduction.....	110
5.2.	Material and methods.....	111
5.2.1.	Biological material.....	111
5.2.2.	Nucleic acid purification.....	111
5.2.3.	Validation by agarose gel electrophoresis.....	112
5.2.4.	Validation by Polymerase Chain reaction.....	112
5.3.	When do the nucleic acids leave the ChipFilter?.....	112
5.4.	What is the minimum number of cells required to obtain nucleic acid by ChipFilter workflow?.....	113
5.5.	How are the nucleic acids retained by the ChipFilter?.....	114

5.6.	Do the reagents used in proteomic sample preparation affect the nucleic acids?.....	116
5.7.	Is the DNA recovered compatible with PCR analysis?.....	118
5.8.	Conclusions.....	119
6.	Toward a multi-omics approach for studying a microbiome standard of the gut	121
6.1	Introduction.....	122
6.2.	Material and methods.....	123
6.2.1.	Biological material.....	123
6.2.2.	Proteomics analysis.....	124
6.2.3.	Nucleic acid analysis.....	125
6.3.	ChipFilter allows the identification of proteins from most of the species of the gut standard.....	126
6.4.	Evaluation of the effectiveness of the preparation according to the microbial groups.....	127
6.5.	Biomass estimation by proteomics.....	130
6.6.	Nucleic acid retrieval.....	133
6.7.	Conclusions.....	135
7.	ChipFilter adapted for aerosol-borne microbial sample collection.....	137
7.1.	Introduction.....	138
7.2.	ChipFilter directly connected to a pump for sample collection...	139
7.2.1.	Experimental design.....	139
7.2.2.	Application.....	140
7.3.	ChipFilter with 18-gauge needle for sample collection.....	142
7.3.1.	Experimental design.....	142
7.3.2.	Application.....	143
7.4.	Conclusions.....	144
8.	Overall conclusions and future perspectives.....	146
	Annex 1 - Description of the ChipFilter Workflows.....	153
	Annex 2 - Label-free quantification intensity for mFASP, in-gel and in-solution workflows for mix 1 and mix 3.....	155
	Annex 3- List of microorganisms used for building database for murine faeces microbiome data analysis.....	159
	Annex 4 – List of microorganisms identified in analysis of murine faeces microbiome using ChipFilter.....	161
	Annex 5 – Article.....	172
	References.....	195

List of Figures

Figure number	Title	Page number
1	The term microbiome comprises both the microbiota (community of microorganisms) and their “theatre of activity” (structural elements, metabolites/signal molecules, and the surrounding environmental conditions)	3
2	Environmental microbiomes that have been extensively studied.....	4
3	Microbiomes are associated in a diversity of disorders, including skin, metabolic, cardiovascular disease, cancer (gastric, colorectal, and oesophageal), infection, and neurodegenerative and psychiatric disorders.....	9
4	Explanation of the methods used in the assessment of the microbiome.....	12
5	A simplified schema in the cultured and uncultured microbiome genomic sequencing.....	14
6	Schema of genomic sequencing with emphasis on targeted and shotgun approaches.....	14
7	Whole genomic sequencing (WGS) pipelines used in metagenomics for taxonomic profiling and functional annotation.....	18
8	Marker/Targeted gene sequencing pipelines used in metagenomics.....	19
9	The challenges at different stages of the metagenomics workflow.....	21
10	Schema for sample preparation workflow for bottom-up proteomics.....	26
11	Schema for different steps in tandem MS/MS.....	29
12	Difference between SRM/MRM, PRM, DDA, and DIA modes used during MS/MS.....	30
13	Different pipelines to overcome the data analysis challenges in metaproteomics.....	31
14	Results that can be obtained after metaproteomics data analysis.....	32
15	Challenges in different stages of the metaproteomics experiment.....	35

16	Representation of the different steps involved in the mould fabrication and PDMS device fabrication.....	37
17	Different microfluidics systems for microbial culture and interaction studies.....	41
18	Application of droplet microfluidics systems for microbial sorting and mutant selection studies.....	42
19	Application of microfluidic channels for microbial detection through surface coating with antibodies, antimicrobial peptides (AMPs), aptamers, carbohydrate-binding molecules (lectin), bacteriophages, and molecularly imprinted polymers (MIPs).....	42
20	Application of different microfluidic devices for microbial collection from bioaerosols.....	44
21	Microfluidic device (ChipFilter) design for bottom-up proteomics sample preparation.....	50
22	Methods for the processing of single cells for multi-omics studies.....	51
23	Multi-omics studies for microbiome.....	53
24	Microbial phylum commonly found in the gut.....	55
25	The overall organisation of the thesis work.....	58
26	Steps discussed in chapter 2.....	61
27	Diagram of two masks necessary to create the mould for ChipFilter.....	63
28	The structures after PDMS mould replication.....	64
29	The key steps involved in sample preparation using ChipFilter.....	64
30	The decrease in contaminants with increasing target cell numbers 10^4 , 10^6 , and 10^7	69
31	The proteins, peptides, and PSMs identified for <i>E. coli</i> , <i>B. subtilis</i> , and <i>S. cerevisiae</i> using ChipFilter.....	71
32	The identified target proteins, distribution of target proteins between each other, and species for ChipFilter WF1 and WF2.....	73
33	The identified total peptides, PSMs, and distribution of peptides between for ChipFilter WF1 and WF2.....	74
34	The physiochemical properties like missed cleavages, peptide length, theoretical positive ion mass, and hydrophobicity for ChipFilter WF1 and WF2.....	75
35	Steps discussed in chapter 3.....	77
36	Steps used in different procedures discussed for microbial sample preparation for LC-MS/MS proteomics.....	81

37	Time taken for seven key steps in proteomics sample preparation by different procedures discussed.....	83
38	Performance assessment of ChipFilter workflow compared to other procedures for Mix 1. The number of identified proteins and peptides. Distribution of the proteins and peptides identified. Distribution of the protein identified according to the species.....	86
39	Performance assessment of ChipFilter workflow compared to other procedures for Mix 3. The number of identified proteins and peptides. Distribution of the proteins and peptides identified. Distribution of the protein identified according to the species.....	88
40	Physical chemical characteristics of the peptides generated by different preparation procedures for mix 1 and mix 3. Positive ion mass (MH+), amino acid length distribution, percentage of missed cleavages, PSMs distribution, and Kyte-Doolittle hydrophobicity index.....	92-93
41	Label-free quantification intensity for ChipFilter workflow for mix 1 and mix 3.....	94
42	Steps discussed in chapter 4.....	98
43	Phylum abundance in different murine faeces samples processed by ChipFilter workflow.....	104
44	Genus's level distribution between different murine faeces samples processed by ChipFilter workflow.....	105
45	Top 20 biological processes seen for the identified proteins from murine gut microbiome.....	106
46	Top 20 cellular localisations seen for the identified proteins from murine gut microbiome.....	107
47	Top 20 molecular functions seen for the identified proteins from murine gut microbiome.....	107
48	Steps discussed in chapter 5.....	110
49	Monitoring the removal of the nucleic acids at different steps of the ChipFilter WF4.....	113
50	Minimum human (10^6) and microbial (mix 1) cell number necessary for visualization in agarose gel electrophoresis using ChipFilter WF4.....	114
51	Representation of the DNA and RNA retention onto ChipFilter.....	116
52	The effect of proteomic reagents on DNA.....	117
53	The effect of proteomic reagents on RNA.....	118

54	PCR products formed from the amplification of plasmid DNA isolated from <i>E. coli</i> with ChipFilter WF4.....	119
55	Steps discussed in chapter 6.....	123
56	Proteins identified for different species present in the Gut standard by ChipFilter WF5.....	127
57	TopN spacing distribution of precursor ions selected for fragmentation by MS/MS as seen in RawMeat version 2.1 for the three replicates of gut standard.....	131
58	Comparison of biomass estimate done by ChipFilter-based metaproteomics with the next-generation sequencing methods and cell number for more than 1% abundance and less than 1% abundance.....	132
59	Nucleic acid integrity obtained after ChipFilter WF5 and DNA isolation kit.....	134
60	Nucleic acid integrity obtained during library generation using TapeStation after ChipFilter WF5 and DNA isolation kit.....	135
61	Steps discussed in chapter 7.....	139
62	Setups to check the stability of the device for aerosol collection with an air compressor and vacuum pump.....	140
63	Representation of the setup to test the stability of ChipFilter using an air compressor pump and vacuum pump.....	141
64	ChipFilter stability analysis shows disruption of the reaction chamber at high injection pressure as compared to the control.....	141
65	ChipFilter modification to accommodate the connector and needle.....	143
66	ChipFilter application in bioreactor sampling.....	144

List of Tables

Table number	Title	Page number
1	A list of different bottom-up proteomics workflows with usage, advantage, and limitation.....	28
2	Comparison of different microfluidic cell lysis methods on efficiency, lysis time, technical difficulty, and cost.....	46
3	Number of proteins identified for each microbial species in mix 1 with every procedure.....	87
4	Number of proteins identified for each microbial species in mix 3 with every procedure.....	89
5	Comparison of overall identifications obtained by different approaches used for murine protein database construction.....	101
6	Number of identified proteins, peptides, PSMs, and unique peptides of four major phyla for four samples processed by ChipFilter workflow.....	103
7	List of microorganisms with their classification and composition in the gut standard.....	124
8	Number of proteins, peptides, and PSMs identified for Gut standard by ChipFilter WF5.....	126

List of Abbreviations

LB	Luria Bertani
PBS	Phosphate buffered saline
IPGG	Institut Pierre Gilles de Gennes
μ	micro
°C	degree Celsius
M	Molar
pH	potential hydrogen
n	nano
UV	Ultraviolet
cm	centimetre
J	Joule
W	Watt
D	Dimension
g	Grams
PDMS	Poly dimethyl siloxane
FASP	Filter-assisted sample preparation
MS	Mass spectrometry
LC	Liquid chromatography
min	Minute
ABC	Ammonium bicarbonate
DTT	Dithiothreitol
IAA	Iodoacetamide
ODG	Octyl β-D-glucopyranoside
WF	Workflow
TFA	Trifluoro Acid
AGC	Automatic gain control
DDA	Data-dependent acquisition
V	Volt
ACN	Acetonitrile
PEEK	Polyether ether ketone
PSMs	Peptide spectral matches
NGS	Next generation sequencing
CRC	Colorectal cancer
IBD	Inflammatory bowel disease
DNA	Deoxyribonucleic acid
PCR	Polymerase chain reaction
CAMI	Critical Assessment of Metagenome Interpretation
WGS	Whole genomic sequencing
MAG	Metagenome assembled genome
TCA	Trichloroacetic acid

SP3	Single pot solid phase enhanced sample preparation
S-trap	Suspension trap
m/z	Mass to charge
MALDI	Matrix-assisted laser desorption/ionization
ESI	Electrospray ionization
DDA	Data-dependent acquisition
DIA	Data independent acquisition
MRM	Multiple reaction monitoring
SRM	Single reaction monitoring
BLAST	Basic local alignment search tool
EST	Expressed sequence tags
PD	Proteome discoverer
PCA	Principal component analysis
MIP	Molecularly imprinted polymers
SDS	Sodium dodecyl sulfate
MALBAC	Multiple annealing and looping-based amplification cycles
SEM	Standard error of the mean
MH+	Positive ion mass
LFQ	Label-free quantification
IBD	Inflammatory bowel disease
HCl	Hydrochloric acid
NADP	Nicotinamide adenine dinucleotide phosphate
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
i.d.	Internal diameter

1. General Introduction

1.1. Microbiome- History and definitions

The term “microbiome” was first coined by Whipps et al (Whipps et al., 1988) and it states,

“A convenient ecological framework in which to examine biocontrol systems is that of the microbiome. This may be defined as a characteristic microbial community occupying a reasonably well-defined habitat that has distinct physio-chemical properties. The term thus not only refers to the microorganisms involved but also encompasses their theatre of activity”.

This definition emphasises both the microorganisms and their interactions with each other and/or with their environment (referred to as theatre of activity). Over the period, several adjustments to the definition were made to fit different themes such as ecology (Marchesi & Ravel, 2015; Orozoco-Mosqueda et al., 2018; Lederberg & Mccray, 2001), organisms/host relationship (Merriam-Webster.com, 2023), genomic/method-driven (Merriam-Webster.com., 2023; Are all et al., 2019; Schlaeppli & Bulgarelli, 2015), or combined (Rogers et al., 2016; Ho et al., 2018; White side et al., 2015). These adjustments have their innate advantages and disadvantages as discussed by Berg et al., 2020 (Berg et al., 2020). Hence, the search for a comprehensive definition encompassing different related themes was coined recently through a panel of international experts discussing the current gaps in the frame of the European-funded Microbiome Support project. According to this panel, two clarifying paragraphs to the definition of Whipps et al. were made highlighting the difference between the terms “microbiome” and “microbiota” by articulating their dynamic character,

- The *microbiome* is defined as a characteristic microbial community occupying a reasonably well-defined habitat that has distinct physio-chemical properties. The microbiome not only refers to the microorganisms involved but also encompasses their theatre of activity, which results in the formation of specific ecological niches. The microbiome, which forms a dynamic and interactive micro-ecosystem prone to change in time and scale, is integrated into macro-ecosystems including eukaryotic hosts, and here is crucial for their functioning and health (Berg et al., 2020).
- The *microbiota* consists of the assembly of microorganisms belonging to different kingdoms (prokaryotes (bacteria, archaea), eukaryotes (algae, protozoa, fungi, etc), while “their theatre of activity” includes microbial structures, metabolites, mobile genetic elements (such as transposons, phage, and viruses), and relic DNA embedded in the environmental conditions of the habitat (Berg et al., 2020).

As shown in figure 1, the most recent definition and usage of the term microbiome will be agreeable to use in studies leading up to the characterization of the

microorganisms along with their internal structural elements like nucleic acids, proteins, lipids, polysaccharides, metabolites (signalling molecules, toxins, and other organic materials), external structural elements including viruses, relic DNA, mobile genetic elements and environmental conditions that influence the entire process.

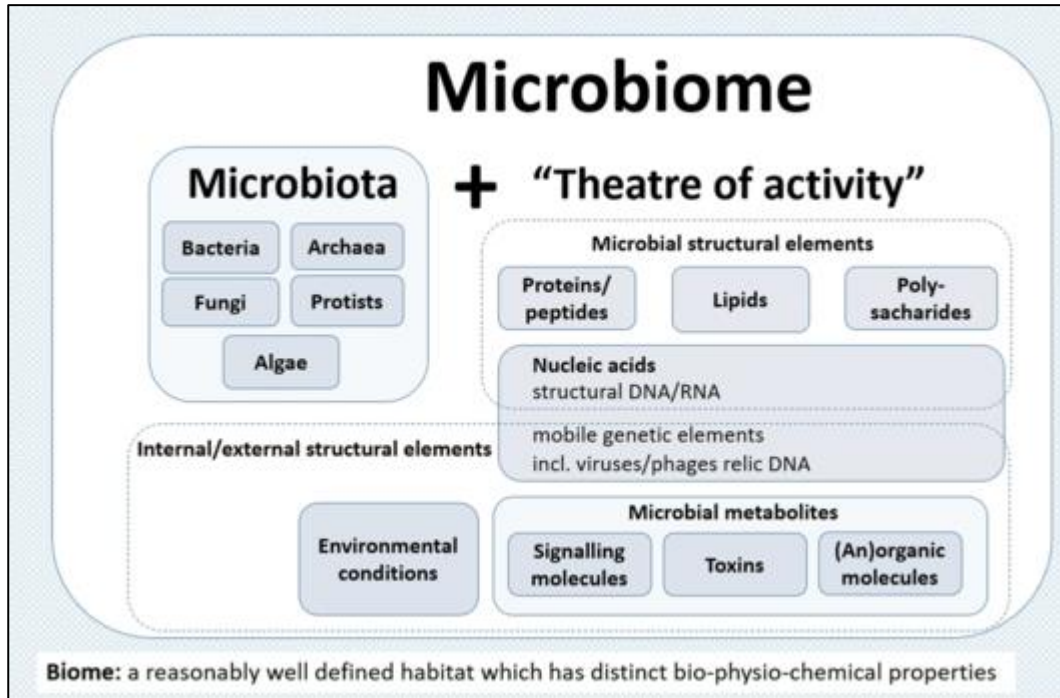


Figure 1: The term microbiome comprises both the microbiota (community of microorganisms) and their “theatre of activity” (structural elements, metabolites/signal molecules, and the surrounding environmental conditions) (Berg et al., 2020)

Studies on microorganisms have been performed since the 16th century with each era witnessing milestone scientific discoveries like the discovery of microorganisms (Lane, 2015), pasteurization (Berche, 2012), antibiotics (Aminov, 2010), the discovery of archaea (Woese & Fox, 1977), biofilm (Flemming, 1993) and bacterial genome sequencing (Fleischmann et al., 1995). These discoveries were possible through technological developments such as microscopy, in-vitro cultivation methods, mass spectrometry, Sanger sequencing, polymerase chain reaction, and next-generation sequencing. The emergence of new technologies and subsequent scientific discoveries has led to a broadening of the definition of the microbiome. Given a better understanding, the different microbiome studies can be separated into environmental (outside a host body) and human/host (inside or on the host body). This classification is based on the probable consequence that the microbiota can have on the human surrounding and health. The

subsequent topics highlight the important projects and discoveries made through environmental and human microbiome research.

1.1.1. Environmental microbiome

Microbiomes of the environment is very diverse and distinct based on their ecosystem. They thrive and steer the activities exclusive to their environment (such as rain cloud formation in the atmosphere (Morris et al., 2010; Georges et al., 2014)). Identifying the type of microorganisms, metabolisms and their interactions is critical for researchers studying environmental microbiomes. Figure 2 highlights the prominent sites in the environment where microbial communities have been studied. This includes marine or freshwater, soil, wastewater or activated sludge, human activity like mining, and the atmosphere.

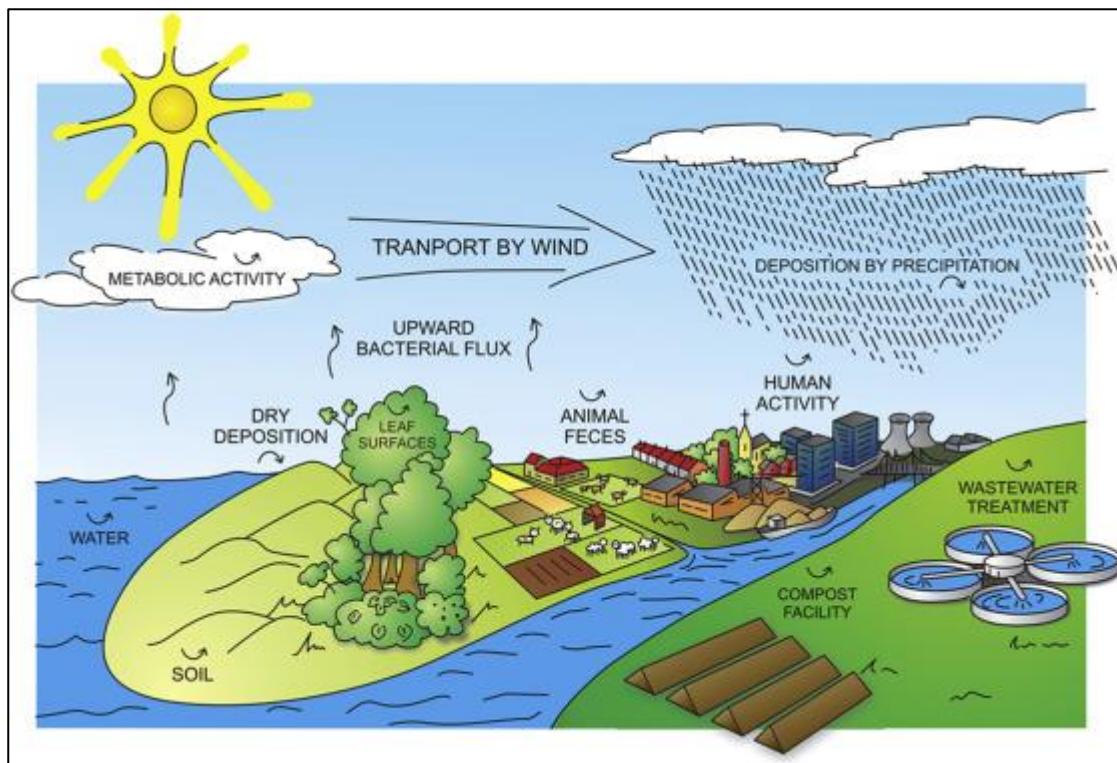


Figure 2: Environmental microbiomes that have been extensively studied (Smets et al., 2016)

Marine or freshwater

The marine habitat is the Earth's largest aquatic ecosystem and houses diverse microbial communities that play a central role in regulating the biogeochemical cycling of carbon, nitrogen, and phosphorus, as well as various micronutrients and trace metals (Azam & Malfatti, 2007). It has been reported that marine microbes adapt to different nutrient environments by expressing a high amount of transporter proteins like

periplasmic binding proteins of ATP-binding cassette transporters and Tripartite ATP-independent periplasmic transporters (Morris et al., 2002; Morris et al., 2010; Georges et al., 2014). These transporters play a very important role in the microbial carbon and nitrogen cycles. In addition, microorganisms have evolved separate metabolic strategies to utilize hydrogen (Kleiner et al., 2012), one-carbon compounds (Sowell et al., 2011), urea (Wang et al., 2011 (B); Alonso-Saez et al., 2012), and taurine (Williams et al., 2012) as energy sources. Trace metals are important nutrients required for bacteria to survive on the Earth, and metalloproteins play vital roles in catalysing necessary biogeochemical reactions (Yong et al., 2014; Cvetkovic et al., 2010). The uptake of limiting metals is a crucial driver of the ongoing adaptive approaches developed by microbes. Low-light intensity, hypoxic and extreme stress environments greatly challenge microbial survival. Proteogenomic research has shown that green sulphur bacteria found in the Antarctic have developed special proteins to aid in higher light capture and counter high sulphur content in the sea (Ng et al., 2010; Lauro et al., 2011).

Soil

Soil covers nearly all the terrestrial areas and hosts the most abundant and diverse microbiota on Earth, making it into another complex and dynamic ecosystem. Soil microbiota participates in the breakdown and conversion of soil materials, contaminant remediation, rhizospheric soils, and semiarid soils, as well as the biogeochemical cycling of carbon, nitrogen, and other biogenic elements (Bastida et al., 2010). Proteins including a wide variety of dehydrogenases, hydrolases, proteases, peptidases, catalases, nitrogen fixation, and nitrification enzymes are identified with the potential to contribute to the biogeochemical cycling of elements as well as in the oxidation of organic matter are found in the semiarid-soils-microbiome (Bastida et al., 2015; Bastida et al., 2014). The microbial decomposition of dead-leaf litter is a critical carbon and nutrient cycling process of terrestrial ecosystems. Extracellular hydrolytic enzymes produced by fungi play a great role in this process and it is regulated by the high nutrient content (Schneider et al., 2012; Becher et al., 2013). Microbial communities in the soil also participate in the remediation and contamination removal as seen in studies performed on uranium-contaminated sites (Wilkins et al., 2009), organic pollutants (Benndorf et al., 2007), cadmium-contaminated sites (Singleton et al., 2003), and hydrocarbon-polluted soil (Bastida et al., 2010; Bastida et al., 2016).

Wastewater or activated sludge

Different microbial communities have been developed to play important roles in the wastewater treatment process including biological phosphorus removal (Albertsen et al., 2016; Skennerton et al., 2015), nitrogen cycling (Speth et al., 2016), metal remediation (Carla et al., 2007), and antibiotic resistance gene study (Raza et al., 2021). The biological

phosphorus removal process using microbial communities has identified differential expression of more than 100 unique proteins from the ATPases, oxidoreductases, and transport proteins family (Wilmes et al., 2008 (A); Wilmes et al., 2008 (B)). The identified proteins closely matched the Candidatus *Accumulibacter phosphatis* (He et al., 2010). In another study, the Candidatus *Competibacter* group showed higher protein activity in the phosphorus removal of activated sludge in continuously stirred tank bioreactors (Bize et al., 2015). Studies on the nitrogen cycling process in the microbiome have led us to the discovery of novel biological processes like partial nitrification-anammox (Speth et al., 2016) and the involvement of bacterial species like *Nitrospira* in complete nitrogen cycling (Daims et al., 2015). Bioremediation of heavy metals including Magnesium, Copper, Zinc, Cadmium, Iron, Chromium, Arsenic, Lead, and Nickel has been partially or completely metabolized by microbial communities (Sharma et al., 2021). Omics studies on these samples have given insight into the novel species and metabolism identification. Finally, the involvement of microbiome in the wastewater to develop antibiotic resistance genes is a topic of concern. It is likely to occur in pharmaceuticals, stock farming, and other ventures (Kim & Carlson, 2007). Metagenomics has been used to elucidate the distribution, horizontal transfer, and degradation-related mechanisms of antibiotic-resistance genes in wastewater treatment plants (Chen & Miller, 2022).

Acid mine discharge

Human activities cause acid mine discharge which is a notable environmental problem. This results in the formation of low pH (acidic) metal-enriched waters that requires treatment before discharge from energy industries and mining activities. Microorganisms associated with this process are important as they have effects on the formation, pollutant release, and biological remediation (Xie et al., 2011). The proteins related to refolding, oxidative stress, and cytochrome central iron oxidation were found in *Leptospirillum* group II species concluding their survival and formation of acid mine discharge in biofilms (Ram et al., 2005; Lo et al., 2007). The different species survive in these conditions through genetic recombination (Denef et al., 2009). A proteogenomic analysis on the *Leptospirillum* group II and III identified from biofilms showed 20% protein identification corresponding to plasmid proteins. Proofs for the upregulation of genes involved in motility, signal transduction, and transport have been identified (Denef et al., 2009; Goltsman et al., 2009). These plasmid proteins correspond to community essential functions like the biosynthesis of vitamins, fatty acids, and biopolymers.

Atmosphere

Atmospheric microbiome characterization is one of the final frontiers yet to be fully conquered. Challenges associated with sample processing limiting the widespread

atmospheric microbiome studies are changes in microbiome distribution, lower biomass, the requirement for longer collection procedures, transportation loss, and the selection of good sample methodology (Dommergue et al., 2019). Several groups have studied the microbial composition and function in the atmosphere for microbial dissemination (Stewart et al., 2021), disease transmission (Moelling et al., 2020), and rain cloud formation (Šantl-Temkiv et al., 2022). The atmosphere acts as the vehicle for the transport of microbial communities, microbial emission from the surface is affected by factors like biofilms, ontogeny, seasonal growth, hydration, particulates, and surfaces. On the other hand, deposition is affected by allometry, ice nucleation proteins, precipitation, and wind speeds. Metagenomic and metatranscriptomics profiling of atmospheric communities has revealed a prevalence of genes encoding cold shock, oxidative stress, and UV repair enzymes (Aalismail et al., 2019; Amato et al., 2019; Archer et al., 2022). Long-range transmission of pathogenic microorganisms including fungi, bacteria, *Plasmodium* parasites, and viruses (SARS and Corona) has been reported (Šantl-Temkiv et al., 2022). Finally, the involvement of ice nucleation responsible for rain cloud formation in the atmosphere is reported in *Pseudomonas syringae* (de Araujo et al., 2019).

1.1.2. Human/Host microbiome

Microorganisms have thrived everywhere on the earth, including within humans. Unlike the common misconception that they can only be harmful, we thrive in cooperation. They assist in different metabolic functions like absorption, digestion, and elimination. In the human body microbial communities are likely to be found in the skin (Grice et al., 2009), conjunctiva (Todar 2006; Cui et al., 2013), oral cavity (Floyd et al., 2010), gastrointestinal tract (gut) (Arumugam et al., 2011), urinary tract (Perez-Carrasco et al., 2021), vagina (France et al., 2022), placenta (Aagaard et al., 2014), uterus (Baker et al., 2018), nasal cavity (Bassis et al., 2014), lung (Moffatt et al., 2017) and biliary tract (Binda et al., 2022). Studies were accomplished by assessing the biofluids like sweat, skin swabs, tears, saliva, faeces, mucous, urine, or blood collected from different donors or through biopsies in specific cases.

Human microbiome project

Recent research in 2016 has shown that the number of human cells and the number of microbial cells in an average normal human body is almost the same (Sender et al., 2016 (B)). A misconception that existed earlier in this finding is that the ratio of the human cell to microbes is 1:10 (Luckey, 1972). So how was this deepening of knowledge made possible? The answer to the question is in the ambitious Human Microbiome Project (2007-2014) and Integrative Human Microbiome Project (2014-2016) (<https://commonfund.nih.gov/hmp/>). Firstly, the Human Microbiome Project has launched to sequence whole genomes of microbial communities isolated from human

donors by culture-independent methods. More than 5,000 samples were collected from tissues, and body sites such as the mouth, nose, skin, lower intestine (stool), and vagina. The project resulted in the characterization of more than 10,000 microbial species that inhabit the human ecosystem. Secondly, Integrative Human Microbiome Project was launched with three sub-projects dealing with conditions related to pregnancy and preterm birth, inflammatory bowel disorder, and type-2 diabetes. Study methods involved 16S rRNA profiling, whole metagenome shotgun sequencing, metatranscriptomics, metabolomics/lipidomics, and immunoproteomics. Combined efforts under this 10-year-long project have led to more than 650 scientific papers that have been cited more than 70,000 times until 2017 (Banquero & Nombela, 2012). The establishment of reference databases for the microorganisms of the human host, clinical and pharmaceutical applications of the studies are some of the key deliverables.

Microbiome as a human organ

A popular hypothesis is to consider the “microbiome” of the human host as an “organ” of the host itself (Banquero & Nombela, 2012). Similarities can be drawn between the microbiome and the organs in several ways such as both being inherited, having physiology and pathology and the individual health can be damaged if the collective structure will be altered. Inheritance of the microbiome is a subject of great interest and is explained by ‘small world’ power law dynamics wherein a core group of bacteria is inherited during birth and this gradually attracts more species towards them during development to enable the formation of communities.

Diagnostics and disease association

Inherently, these microbial communities have been identified to be directly or indirectly associated with diseases or disorders as shown in figure 3. The list includes bacterial, fungal, and viral infections that are often caused by single cells or closely related microbial species. Disorders of the liver, gastrointestinal tract, metabolism, cardiovascular, and neurology. Evidence has surfaced for the association of microbiome in psychiatric diseases as well, although it requires exhaustive studies (Chen et al., 2021). Finally, the association between cancer and microbiome is available.

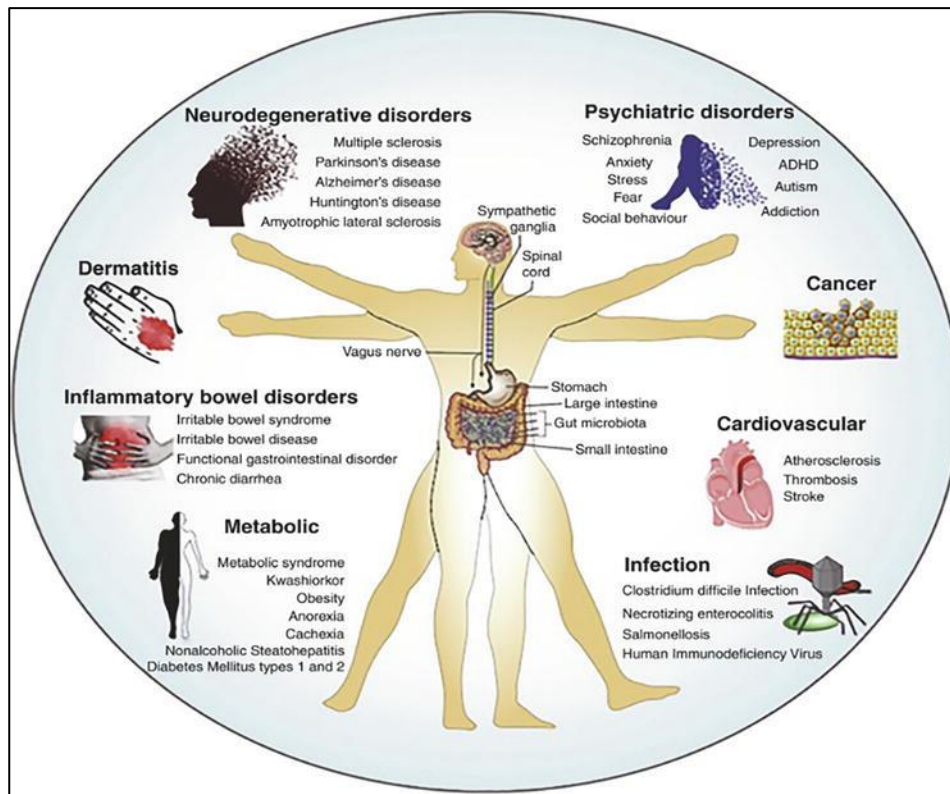


Figure 3: Microbiomes are associated in a diversity of disorders, including skin, metabolic, cardiovascular disease, cancer (gastric, colorectal, and oesophageal), infection, and neurodegenerative and psychiatric disorders (Goodman, 2021)

Microbiome-based diagnosis of disorders is a very attractive option, it will be certainly non-invasive and can utilize the existing molecular diagnostic tools like polymerase chain reactions or next-generation sequencing (NGS) for nucleic acids, and mass spectrometers (MS) for proteins or metabolites. The possibilities and potential are very high, but the bottleneck is in sample collection and processing, data analysis, and ethics or quality control (Schlaberg, 2020). Several research organizations and companies (Luis, 2021) have been working to develop a solution to this problem. The current research work will as well provide a solution that will allow easy collection, transportation, and processing of biofluids.

Prognosis and therapeutics

Current medical interventions have been focused on improving the microbiome distribution to fight against a specific disease as well as identifying biomarkers within the microbiome. Diseases associated with the microbiome include Inflammatory bowel disease (IBD), diabetes, infections, colorectal cancer (CRC), neurological conditions (Depression, Alzheimer's disease), and metabolic disorders (Baohong et al., 2017; Fan & Pedersen, 2021; Ping et al., 2017). Microbiome therapeutics are aimed at engineering the

gut microbiome using additive, subtractive, or modulatory therapy with an application of native or engineered microbes, symbiotic, bacteriophages, and bacteriocins (Yadav et al., 2021). Additive therapy involves the use of probiotics or faecal microbiota transplant. Subtractive therapy purposes to reduce the deleterious pathogens from the microbiome through the use of the antimicrobial activity of bacteriocins (ribosomally synthesized peptides exhibiting antimicrobial activity) and bacteriophages without the use of antibiotics. Modulatory therapy involves the usage of diet, exercise, or antibiotics to restore or modify the microbiome in the human body. Psychobiotics are a group of agents that have a psychotropic effect on anxiety, depression, and stress that may be probiotic (live microbes), postbiotic (inanimate microbial components), prebiotic (stimulants for colon microbial growth), or symbiotic (mixture of pro- and pre- biotics) and target the gut–brain axis and confer mental health (Dinan et al., 2013). Attractive options for microbiome therapeutics exist with inborn challenges in safety, stability, microbial characterization, disease specificity, ethics, and suitability.

Nutrition

The nutritional intake significantly regulates the microbial composition of the different communities in the human body (Valdes et al., 2018). Direct modulation of the gut microbiome can be witnessed by the usage of probiotics or prebiotics. Research is being carried out for dietary-based modulation of the microbiome as it is a very important field operating on the effect of environmental factors (food and lifestyle) decided by the host on its microbiome.

1.1.3. Approaches in microbiome studies

In this part, we have simplified our understanding of the various microbiome based on the earth into two groups – environmental and host/human. Even though the groups are different the microbial structural elements are the same. Hence, the technologies available for studying these biomolecules are equally available for both groups. However, the difference will arise in the prior stage where the microbial cells constituting the community interact (such as host cells, chemicals from the environment, physical factors like pH or salinity, etc) leading to their existence. These internal and external structural factors decide the complexity of studying a specific microbiome.

The selection of the approach is determined by the questions that one expects to answer. This is shown in figure 4. The three broad questions that can currently be answered by available technologies are (i) who is there? (ii) what can they do? (iii) what are they doing? In many cases more than one of the questions can be answered with limited sensitivity or an in-depth answer to a single question can be achieved by integrating more than one technology.

Who is there?

The most fundamental question in microbiome research is to understand the list of species present in the community. The depth of understanding ranges from knowing the phenotype that can be achieved by simple staining approaches such as Gram's staining and visualising the cells under the microscope to taxonomic classification by genomic sequencing. Phenotypic information can be achieved by cell culture methods for most microorganisms and they are usually restricted to single-cell to a few hundred cell types. In communities with higher diversity and limited culture possibilities, genomic sequencing by shotgun sequencing is done which is also known as metagenomics. Whole cells or DNA material are usually used to identify the taxonomy and species diversity. Greater success in species identification can be achieved by utilizing the barcoding approach during genome sequencing. In the case of metagenomics, the reliability of bioinformatic approaches is essential to achieve taxonomic classification. A process known as binning is required to associate a particular genomics sequence to a species (Kunin et al., 2008). Taxonomic classification and phenotypic information detailing the microbial potential can be achieved as an answer.

What can they do?

The metabolic potential of the microbial candidates in the community is crucial to understand the types of benefits or detriment the individuals can have on each other and the community in general. This can be achieved by gene prediction or by whole gene expression studies using metatranscriptomics. The nucleic acid- DNA or RNA are the key components that are sequenced by NGS methods to obtain this information. DNA sequencing can provide the identification of genes that confer metabolic function and thereby predict metabolic function. RNA sequencing can directly provide information about the active genes under a given physiological condition and also help to understand the changes in expression levels during changing conditions. This information can be also used to computationally predict the metabolic pathways active in the microorganism by using systems biology approaches.

What are they doing?

The understanding of active metabolic pathways in the microorganisms will explain their function within the community, thereby establishing the interactions they have with each other and their environment. Microbiomes are dynamic systems and to have sight of the modes of operation of individual players is very useful. The activity of the microbes is recognized by directly identifying the metabolites or proteins that are involved and is done through metabolomics and metaproteomics respectively through MS. Proteins catalyse biological reactions for the survival of the species where the

metabolites are consumed or released. The nature of these reactions is supervised by the substrates offered by the environment. To summarise, we have seen the growing importance of the study of microbiomes in a wide range of areas, from understanding our environment to improving our health. This study is complex because of the extraordinary diversity of microorganisms and the dynamic nature of their interactions. Depending on the question posed, complementary approaches will be favoured, including the metagenomic and metaproteomic approaches that we will present in the following sections.

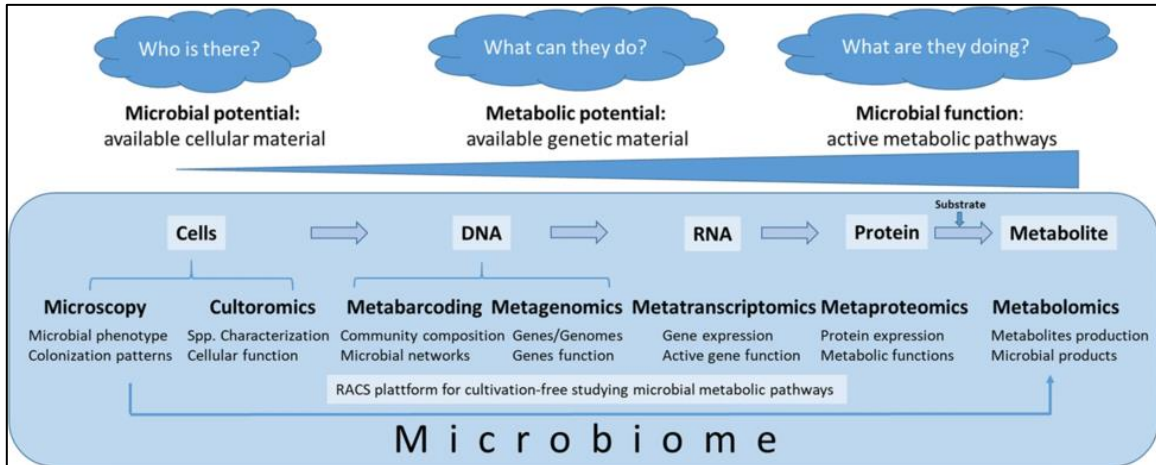


Figure 4: Explanation of the methods used in the assessment of the microbiome (Berg et al., 2020)

To summarise, the study of microbiomes is growing in importance in a wide range of areas, from understanding our environment to improving our health. This study is complex because of the extraordinary diversity of microorganisms and the dynamic nature of their interactions. Depending on the question posed, complementary approaches will be favoured, including the metagenomic and metaproteomic approaches that are presented in the following sections.

1.2. Metagenomics

Metagenomics can be defined as the study of genetic material recovered directly from communities found in environmental or clinical samples. After isolation from the matrix or cells, the genetic materials (genomic DNA, plasmids, or small DNA fragments) are typically sequenced by high-throughput shotgun sequencing methods. The sequences generated are utilized to construct metagenome-assembled genomes (MAGs). The genomes of cultured microbial communities isolated from the environment or clinical isolates can also be sequenced similarly to identify perturbations in the system. The difficulty in such an approach is the hardship of culturing all the isolated species. Hence, researchers propelled toward uncultured metagenomics. A simple outline of the process is shown in figure 5 which describes the steps for metagenomics from environment samples by cellular isolation (red arrows) or cell culture methods (blue arrows). Upon DNA extraction from cell isolation, the sequencing is performed by NGS methods, and the assembling of genomes for the participating species is done.

Metatranscriptomics is a closely associated method that involves the sequencing of RNA to identify the gene expression patterns showing the active genes within a community, to quantify their expression levels, and to monitor how these levels change in different conditions (e.g., physiological vs. pathological conditions in an organism). It provides the advantage to differentiate the active functions of microbial communities which appear to be the same in terms of microbe composition (Bashiardes et al., 2016). Despite the importance of this approach, we will not discuss it here as we did not work on RNA during this study.

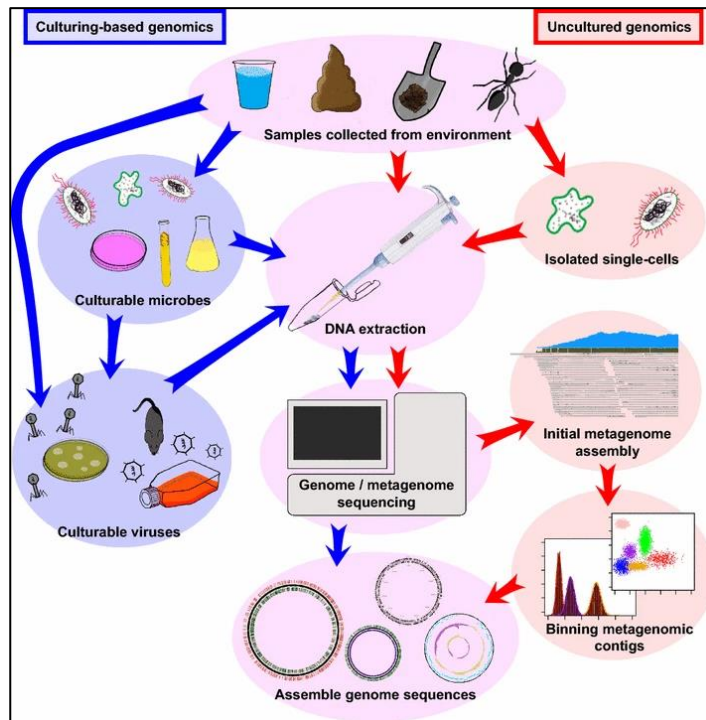


Figure 5: A simplified schema in the cultured and uncultured microbiome genomic sequencing (Thomas et al., 2012)

1.2.1. Steps involved in nucleic acid sequencing

The workflow for metagenomics can be broadly classified into the following steps- (i) Sample collection, (ii) Nucleic acid isolation and library preparation, (iii) Sequencing (iv) Assembly and annotation. The representation of various steps is shown in figure 6.

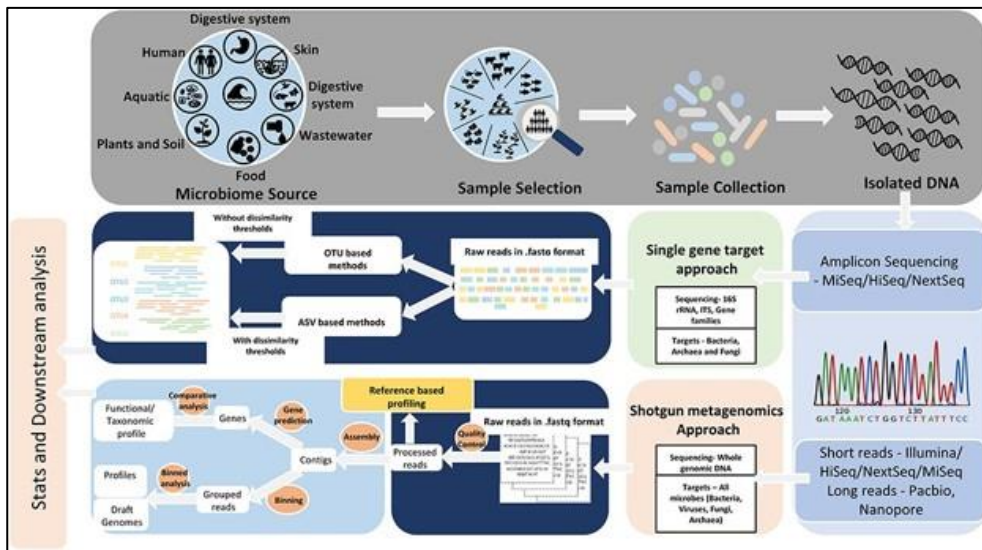


Figure 6: Schema of genomic sequencing with emphasis on targeted and shotgun approaches (Bharti & Grimm, 2021)

Sample collection

The first step in the metagenomics workflow is sample collection. Depending on the source for sample collection, the composition and amount of nucleic acid can be varied. The isolation of the microorganisms from the matrix (environmental) or host tissues is a specific challenge in metagenomics. Based on the material on which the cells bound the complexity may increase such as in the case of soil samples (Lombard et al., 2011) where the steps to isolate the microbiome are significantly higher. Alternatively, in the case of a sample from the atmosphere (Behzad et al., 2015) the cells are more dispersed and hence longer time and volume will be required to isolate sufficient nucleic acid.

Contamination arising during sample collection is a major problem and it can be due to temperature, humidity, or other factors that can alter or contaminate samples (Thomas et al., 2012). Cross-contamination arising due to the closeness of different samples can also be caused. The development of suitable methods for collection to reduce contamination issues is critical.

Transportation of the samples from the site of collection to the site of processing can induce differences in the quality of the nucleic acid extracted as the microorganisms can be unstable during the journey conditions. Thus, it is necessary to include good storage conditions that will include avoiding inconsistent freeze-thaw cycles, usage of chemical preservatives, or usage of on-site sample analysis/preparation (Dominianni et al., 2014; Choo et al., 2015).

Storage of the materials has been shown to influence the output too (Mahalanabis et al., 2009). Positive results in terms of conservation of microbial diversity were realised in the human faecal sample stored rapidly upon collection at -80 °C over storing at 4 °C (Choo et al., 2015). Another important factor to consider is the safety aspect associated with the sample collection process. Since the microorganisms can pose a health threat to the handler, suitable precautions need to be undertaken to ensure safety.

Nucleic acid isolation and library preparation

The nucleic acid (DNA/RNA) isolation method is required to incorporate features like extraction from all types of microorganisms, high quality, high quantity, and minimal biases. For example, the case of bacteria species that are Gram-positive, which contain a peptidoglycan cell wall is usually harder to penetrate (Mahalanabis et al., 2009). The first step in the pre-treatment process is to enrich the cells from their matrix and enable cell lysis. Two main cell lysis methodologies are based on physical/mechanical means (sonication, bead beating, or heating) or by chemical means (chaotropic, detergents, or enzymes) (Psifidi et al., 2015). Often a combination of methods is used to ensure the best

results. The selection of the pre-treatment method is influenced by the microbiome and downstream steps. After cell lysis, enrichment of the nucleic acids by the removal of other cellular components is done by solvent extraction or with silica columns to ensure the binding of only nucleic acids. This is followed by elution of the nucleic acids and reconstitution.

Library preparation is the stage to select the nucleic sequences that will undergo sequencing and indexing to improve downstream quality control. The steps will necessarily include DNA fragmentation or site-specific amplification, then sequence binding to a matrix like beads or transposons (end-repairing) followed by barcoding with adapter sequences. Two types of library preparation methods for shotgun whole genome or targeted amplicon (marker gene) based exist. Shotgun sequences are performed on the entire DNA sequences whereas the targeted amplicon base is usually focused on the rRNA gene region in bacteria (16S) or fungal (internal transcribed spacer). It is to be noted that this targeted amplicon sequencing can include any region of choice. Another important consideration for library preparation in metagenomics is the starting DNA requirement. Microbial cells on average contain few femtograms of DNA, if the number of cells is very low (e.g., 1000 cells) then the DNA extracted will be only in the picogram range (Hutchison & Venter, 2006). Library preparation methods should allow for the incorporation of even this low amount of DNA. The most used library preparation products are commercially manufactured by Illumina, Inc. Three of their products are Nextera™ XT, Illumina® DNA prep (M) Tagmentation, and Illumina® DNA PCR-free Tagmentation. Nextera™ XT, and Illumina® DNA prep (M) Tagmentation require at least 1ng of starting DNA, PCR-based amplification, and uses transposons and bead-linked Tagmentation respectively. Illumina® DNA PCR-free Tagmentation requires at least 25ng of starting DNA, PCR-free method and uses bead-linked tagmentation. In metagenomic library preparation, two approaches (positive and negative) have been developed to reduce the noise arising from host nucleic acids (Chiu & Miller, 2019; Heather & Chain, 2016).

Sequencing

Sequencing is achieved using instruments called NGS platforms. Ideally post-library preparation, the sequencers synthesize complementary strands by addition of nucleotides and simultaneously sequence the spatially segregated amplified DNA templates in a massively parallel fashion without physical separation. The steps in any platform are the same but the operation strategy is unique based on the manufacturer (Anderson & Schrijver, 2010). This also groups the platforms into second and third-generation sequencers (Maljkovic Berry et al., 2019). In second-generation platforms, the generation of template DNA can be achieved by emulsion PCR, rolling circle, and solid-

phase amplification that are dependent on the single DNA molecules, whereas, the third-generation sequencers use single DNA templates directly. For sequencing, approaches like pyrosequencing, reversible terminator chemistry, and ligation-mediated by ligases are utilized in second-generation sequencers and while phospholinked fluorescent nucleotides or Real-time sequencing like nanopore are utilized by third-generation sequencers. Several studies comparing the performance of different platforms for sequencing metagenomic samples for specific analysis have been done (Frey et al., 2014; Miyamoto et al., 2014). Second-generation platforms have greater coverage in terms of length and depth, whereas third-generation platforms offered longer sequence reads and high speed.

Assembly and annotation

Sequenced reads obtained consist of nucleic acid sequences that need to be processed by several steps before obtaining biologically relevant information like taxonomy or functional annotation. The pipeline for different analyses is distinct and some of the most used workflows and software are listed in figures 7 and 8. It must be remembered that more than one pipeline can be built to analyse the data obtained after sequencing and the selection of suitable components in the pipeline is critical to the success of the entire process. The curation of these tools and pipelines for the metagenomics context has been done in global projects like Critical Assessment of Metagenome Interpretation (CAMI) and CAMI II (19, 20).

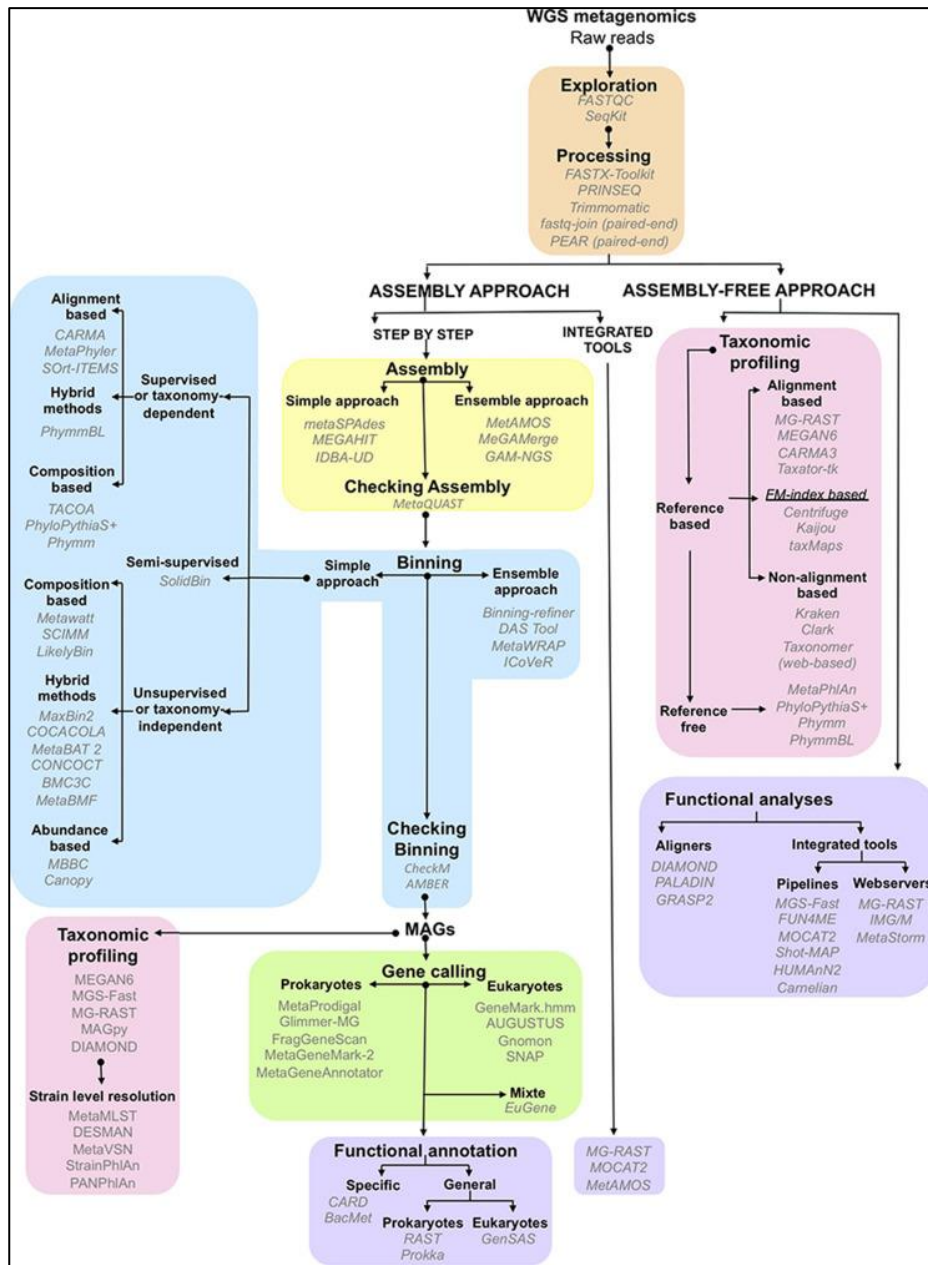


Figure 7: Whole genomic sequencing (WGS) pipelines used in metagenomics for taxonomic profiling and functional annotation (Pérez-Cobas et al., 2020)

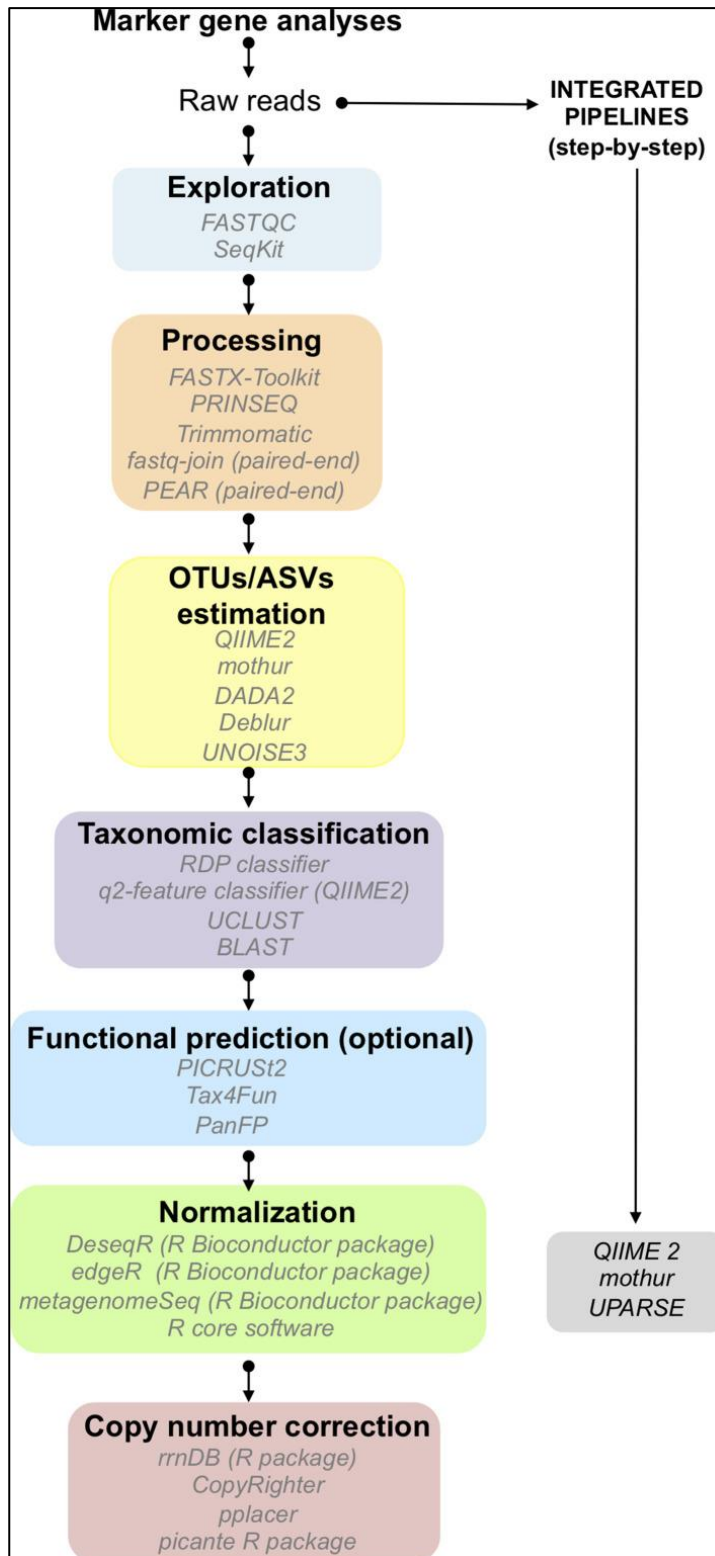


Figure 8: Marker/Targeted gene sequencing pipelines used in metagenomics (Pérez-Cobas et al., 2020)

1.2.2. Accomplishments

“Metagenomics” keyword search in PubMed resulted in 30 119 results from 1998 to the present (13-11-2022). A significant portion of these has been submitted in the last 10 years (2012-2022); indicating that the metagenomics era is still peaking (<https://pubmed.ncbi.nlm.nih.gov/?term=metagenomics&timeline=expanded>).

Metagenomics is revolutionary as it has replaced the need for developing culture-based methods to assess the taxonomy or functional information of microorganisms. Recent developments in sequencing platforms have also made this process extremely fast and portable.

Environmental metagenomics has allowed for a deeper understanding of the participation of different microbial communities in the release or degradation of pollutants, rain formation, and thriving of ecosystems under the ocean or in extreme places. Metagenomics has led to the taxonomic classification of many new bacteria, fungi, archaea, and viruses isolated from different ecosystems. Further, the functional characterization of the genes associated with useful certain pathways has led bioprocess scientists to isolate them and utilize them to produce useful bioproducts (e.g., biofuels (Sartaj et al., 2022))

In the pharmaceutical and food industries, it is now possible to manufacture a range of products with better quality due to the development of metagenomics, including novel enzymes exhibiting thermal stability, and genes conferring antimicrobial resistance (Coughlan et al., 2015). Screening for these genes was made originally from microorganisms isolated from different environments or by plasmid clonal selection from cultured microorganisms. Identification of genes responsible to produce these useful products has allowed the use of recombinant DNA technology to prepare and amplify clones of the gene expressing the production of the bioproducts like insulin.

Using metagenomics, it is now possible to screen samples for pathogens, identify viral genomes, and assess antibiotic resistance and human host response to the microbiome in a clinical setup. This also has led to the growth of personalized medicine and now individuals are willing to perform their gut microbiome sequencing for diagnosis or enrichment. Several vendors offer screening and medical consultations to their customers like in the case of mybioma (<https://www.mybioma.com/en/>). The growth of metagenomics from a scientific domain to offer a deeper understanding of the microscopic world to offer personalized point-of-care is a big accomplishment by itself.

Finally, the biggest contribution of metagenomics is in the development of MAG which will be very useful for the identification of protein sequences in metaproteomics. Metaproteomics cannot be successfully performed without the same population being

sequenced for its genomes. Hence, metagenomics led to the growth of other associated omics approaches like metaproteomics.

1.2.3. Challenges

Metagenomics has developed significantly in recent years, yet the challenges at every stage of the workflow are evident. Figure 9 highlights some of them seen at different stages from study design until data analysis. They are explained in detail in the following sections. Each of these challenges offers opportunities for future research to develop and contribute to the field of metagenomics.

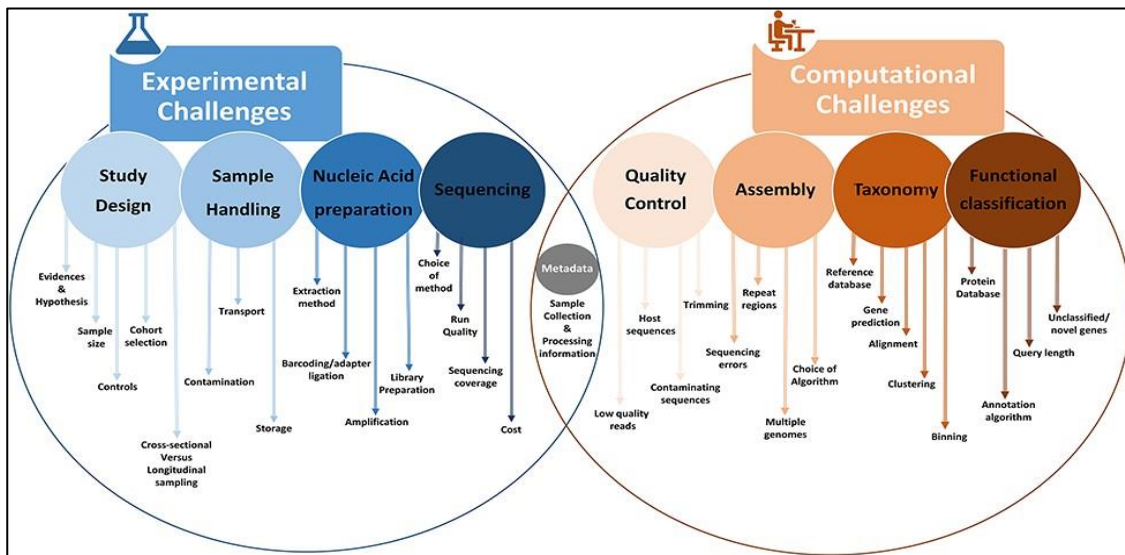


Figure 9: The challenges at different stages of metagenomics workflow (Bharti & Grimm, 2021)

Experimental Challenges

The development of suitable hypotheses and evidence before the experimentation is essential in metagenomics. Metagenomics study provides more open-ended questions than answers, hence it is necessary to identify the scientific questions that can be addressed through the data analysis pipeline to be implemented. Before the actual study, pilot experiments can help in improving data processing and eliminate confounding effects (Honaas et al., 2016). Next, it is critical to include a sufficient sample size as the microbial load can vary between biological replicates (Vandeputte et al., 2017). The inclusion of necessary controls can help to identify contaminants and false positives. But, several conditions including age, gender, ethnicity, diet, genotype, and several other lifestyle factors that can influence the study may not have suitable controls. It is seen in animal studies that genetically identical murine can exhibit different bacterial profiles (Lundeberg et al., 2017; Robertson et al., 2019), variations occur in co-housed animals

and other factors such as strains, housing facilities, and breeding can cause variations (Laukens et al., 2016). Similar challenges exist for environmental samples as well.

Sample handling is a very important step to avoid contamination, loss of materials, or biases. Depending on the community targeted for the study, the sample collection procedure will vary and cautious sample collection will ensure reliable and repeatable results. Usage of special filters for sample collection must be ensuring unbiased collection and minimal interference. Also, it is essential to ensure good transportation conditions and further storage conditions before processing or sequencing.

Nucleic acid extraction is the most critical stage in sample preparation as the results are directly dependent on its quality and quantity. Several challenges include nucleic extraction from low cellular samples, cell lysis challenges, dissociation of matrix materials from the cells, and stability of the nucleic acids. Conventional methods used for culture-based cells may not be always useful for metasamples and thus newer methods need to be developed.

Sequencing platform choice can be dependent on factors like cost, depth, and speed. The development of long-read sequencing platforms (third generation) has opened very attractive options for metagenomics, but the utilization has not been completely reliable due to their dependence on the quality of the nucleic acid material. On the other hand, the polymerase enzymes that amplify short nucleic acid fragments in sequencing methods are error-prone at the terminal.

Computational challenges

Computational challenges are due to the complexity generated by the biological data, lack of metadata information, scarcity of standard formats, and computational resources for high-volume data (Treangen & Salzberg, 2011; Fricke & Rasko, 2014). At first, sequences are subjected to quality control to remove contaminants, trim the adapter sequences, and filter low-quality sequences. The development of these quality control steps is essential. Assembling complex multiple genomes from different species like bacteria, fungi, viruses, and archaea can include sequencing errors, the presence of intergenomic and intragenomic repeats, and uneven sequencing coverage (Abram, 2015; Howe et al., 2014). Further, the assembly process in metagenomics can be complex due to the uneven abundance of multiple genomes in samples originating from the same conditions (Marbouty & Koszul, 2015).

In conclusion, the metagenomic approaches had to face many challenges (figure 9). Among the experimental challenges we have just described, we will see that the ChipFilter microfluidic system can be a solution for the preparation and containment of samples.

Metagenomic approaches have enabled us to better understand complex microbial ecosystems. answering the questions "who is there" and "what can they do"? To answer the question "what are they doing?", complementary omics approaches like metatranscriptomics and metaproteomics. Metatranscriptomics directly correlates taxonomic signatures with profiling of mRNA transcripts generated, whereas metaproteomics associates microbial protein profile.

1.3. Metaproteomics

Metaproteomics can be defined as the large-scale characterization of the entire protein complement of environmental microbiota at a given point in time (Wilmes et al., 2006). This definition was coined in 2006 and since the growth of research on this topic has raised exponentially. Developments in mass spectrometry and separation techniques like liquid chromatography have contributed significantly to this increase. Metaproteomics is closely related to metagenomics and often used as a complementary technique to characterize microbiomes. This is the reason it is often referred to as proteogenomics.

The most imperative understanding of the significance of metaproteomics can be obtained from the questions that the study can provide an answer to with high confidence. The questions are listed below with references from published studies.

Who is there and what abundance?

Microbial community composition can be assessed by quantifying cell numbers or by quantifying biomass for individual populations. Metagenomic techniques have been widely utilized for quantifying cell numbers, but recently biomass quantification has been successfully demonstrated by using metaproteomics approaches to microbiome from two alkaline soda lakes, and saliva (Kleiner et al., 2017). A similar approach will be undertaken in the present study to estimate the biomass from a microbial mixture of 17 species.

What is their expressed metabolism and physiology?

The metabolic state of the microbes in an environment is a key question that can be explained by the active protein or metabolite composition. Metaproteomics delivers the information required for understanding the metabolisms and physiology of the host and microbiota in connection to the host (Kleiner et al., 2012). *Olavius Algarvensis* was shown to have previously unknown energy metabolism shared with its bacterial symbionts to cope with nutrient-limited environments.

How do community members interact?

Upon understanding the metabolic pathways in participating species, it is critical to know how the behaviour of one influence the survival of the others. This relationship can be symbiotic, mutualistic, parasitic, or commensal. Metaproteomics has been used to discover relationships between species that were previously unknown (Hamann et al., 2016). *Arcobacter* and *Breviatea (Lenisia limosa)* were demonstrated to share a mutualistic benefit through differential proteomics that reveals the presence of *Lenisia* stimulates the expression of known 'virulence' factors in *Arcobacter*.

Who uses a specific substrate?

Another interesting study that can be performed using metaproteomics is to determine the adaptation for carbon or nitrogen substrates by different species in a community. The information can be vital to understand the substrate utilization preferences and resource partitioning structure among microbial communities. Environmental metaproteomics has been able to determine this in several ecological hotspots using substrate isotope labelling techniques before proteomics (Jehmlich et al., 2016; Samuel et al., 2016). Proteomic stable isotope probing allows quantitative comparisons of dissolved free amino acid assimilation between sympatric populations and protein functional groups within discrete populations. This allows an unprecedented examination of population-level metabolic responses to resource acquisition in complex microbial communities like coastal marine Bacterioplankton.

What are the carbon sources and assimilation pathways of microbes in the environment?

Another method to identify the carbon sources and assimilation pathways in microbes is using the ($\delta^{13}\text{C}$) measurement by mass spectrometry. The analysis is also known as stable isotope fingerprints and is very useful in environmental microbiology to understand the biogeochemical pathways of microbial communities without labelling (Kleiner et al., 2018).

Further, additional questions like (i) Who takes up, metabolizes, and transfers isotopically labelled substrates? (ii) What is the growth rate of individual community members? (iii) Which proteins exhibit a specifically targeted activity? (iv) What are the abundances of environmental viruses? (v) What are the ages and roles of the cell-free “relic” proteins in the environment? (Kleiner, 2019) needs to be addressed.

1.3.1. Shotgun proteomics

Shotgun proteomics is the most used strategy for metaproteomics. It incorporates the workflow used for bottom-up proteomics with liquid chromatography and mass spectrometry. The three stages of bottom-up proteomics include (i) pre-analytical sample preparation, (ii) analysis of peptides by tandem mass spectrometry, and (iii) data analysis.

Pre-analytical sample preparation

Sample preparation of bottom-up proteomics requires the treatment of cells and proteins to break them into peptides that can be fractionated and analysed by mass spectrometry. Several steps combining physical, chemical, and enzymatic reactions need to be orchestrated to have peptides for MS analysis. The outline of the most vital steps/chemical reactions performed in bottom-up proteomics is highlighted in figure 10. The steps in blue highlight the pre-analytical sample preparation steps that will be later

shown to be performed in the ChipFilter, green is the analytical step and grey indicate the sample to be collected and processed. The complexity of the sample collection and processing varies depending on the source as seen earlier for metagenomics and the choice of workflow and steps involved will greatly vary.

Sample preparation begins with the sample collection of intracellular and extracellular proteins. Unlike nucleic acids, proteins obtained extracellularly are secreted by the microbes or host and are useful to understand the cell-to-cell interactions and enzymatic catalysis with the matrix (Zhang et al., 2018). Efforts during sample collection are focussed to obtain extracellular proteins or peptides along with cells that secrete them to have a comprehensive understanding.

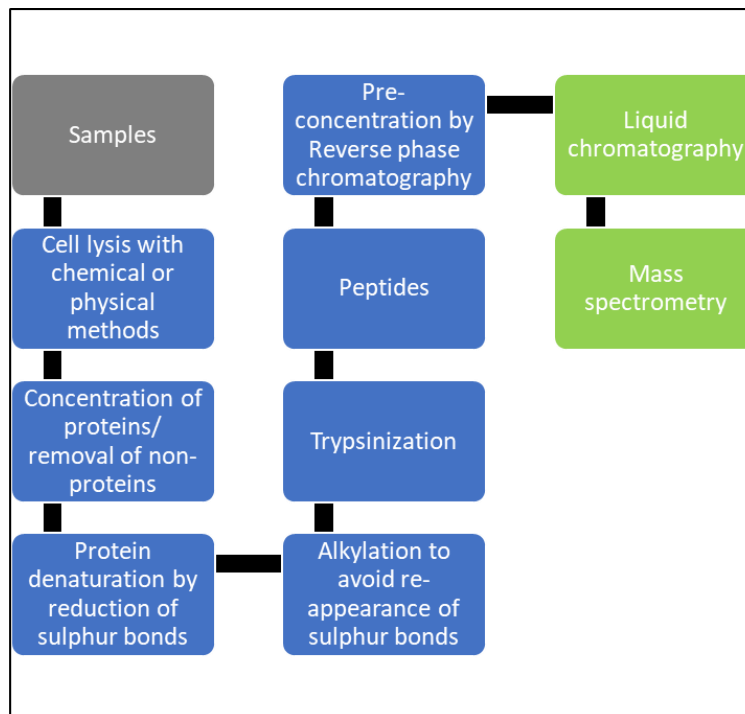


Figure 10: Schema for sample preparation workflow for bottom-up proteomics

To obtain the intercellular proteins, it is necessary to lyse the cells to release the components. Cell lysis can be facilitated by chemical (alkaline/acidic solutions or surfactants), physical (heat, shear force, sonication, radiation), enzymatic (cellulase, lysozyme), or a combination of the methods to obtain high-quality cell lysis and protein solubilization. The choice of cell lysis method, protein pre-concentration, and digestion method are the three critical processes. Several workflows have been established around these processes to enable efficient sample preparation (Varnavides et al., 2022), but the existence of a single universal protocol to perform sample preparation from cell lysis of several microbial species (fungi, bacteria, archaea, and virus) to proteolysis is still

unavailable. In this regard, the present work discusses a novel workflow for metaproteomics utilizing the principles of microfluidics.

Upon cell lysis, the proteins must be separated or enriched from other biomolecules that will interfere with proteolysis later. This can be achieved by protein precipitation, separation by electrophoresis, affinity-based methods, or size exclusion. Protein precipitation is achieved by causing hydrophobic aggregation by disrupting the folded protein structure and exposing the hydrophobic interior or through dehydration of the outer cover in folded proteins. The most common protein precipitation agents used for proteomics are acetone, methanol-chloroform, and trichloroacetic acid (TCA) (Pérez-Rodríguez et al., 2020). Alternatively, aggregation can be achieved using magnetic beads reversibly binding proteins in tubes known as single-pot solid phase enhanced sample preparation (SP3) (Hughes et al., 2019). Protein separation through gel electrophoresis is a classical protocol that has been extensively used. It has the advantage to also fractionate the proteome based on the molecular weight which will add more dimensions to the mass spectrometry analysis. Recent developments have allowed the separation of proteins using molecular weight cut-off nitrocellulose membrane known as filter-assisted sample preparation (FASP) (Wiśniewski et al., 2009) or with porous packed beds known as suspension-trap (S-trap) (HaileMariam et al., 2018). Both technologies utilize gravitational forces for the removal of non-protein substances.

Other reactions include reduction and alkylation before proteolysis. The reduction of cysteine residues allows the denaturation of the protein's tertiary structure which enables better proteolytic activity. Reduction is performed with chemicals like dithiothreitol, tris-(2-carboxyethyl)-phosphine, and mercaptoethanol. Alkylation on the other hand prevents the reformation of the sulphur linkages by alkylating the sites. Alkylation is performed with chemicals like iodoacetamide, iodoacetic acid, acrylamide, and chloroacetamide (Müller & Winter, 2017).

The final step is protein digestion which is necessary to generate peptides. It is achieved using proteolytic enzymes such as trypsin that cleave the peptide bonds at the carboxy-terminal of lysine and arginine. Digestion of proteins can be performed in-solution, in-gel, and on hybrid systems such as FASP, S-trap, or by using magnetic beads binding proteins reversibly as in SP3. Optional steps like peptide clean-up with C18 reversed-phase systems can be performed to remove contaminants before MS. Some of the well-known technologies for sample preparation or protein isolation are mentioned in table 1 with their uses, advantages, and limitations. Among the methods, most of them are useful to process only isolated proteins for MS or can only separate the different biomolecules like nucleic acids or proteins for further processing. Often, they are also

limited by the low quantity of starting material. But, the ChipFilter method discussed in the present work will overcome these challenges.

Product Name	Manufacturer	Usage	Advantage	Limitation
Microfluidic Device	Device proposed in the project	Cell Lysis; Protein processing for MS; DNA and RNA separation	Automation; Useful for Multiomics; Performs all steps for MS sample preparation	Recent technology
Gel based	ThermoFischer	Protein separation and subsequent processing for MS	Classical method; offers fractionation of proteins	Time and labour intensive; No Automation; No cell lysis
In-stage Tips	PreOmics	Protein processing for MS	Automation; robust in cell type and volume	Recent technology; No Multiomics
Magnetic Beads (SP3)	Sigma-Aldrich	Cell Lysis; Protein processing for MS	Automation; robust in cell type and volume	Complexity to handle multiomics; Recent technology
Size exclusion Filters (FASP)	Millipore	Protein processing for MS	Useful for Multiomics; Performs all steps for MS sample preparation	Difficulty to handle low protein sample
Suspension trapping (S-Trap)	Protifi	Protein processing for MS	Useful for Multiomics; Performs all steps for MS sample preparation	Difficulty to handle low protein sample
Sample Preparation Kit	Biognosys	Cell Lysis; Protein processing for MS	Specific for proteomics sample preparation	No Multiomics
EasyPep MS Sample Prep kits	ThermoFischer	Cell Lysis; DNA, RNA and Protein isolation	Separation of all the three biomolecules	No sample preparation for any omics study
TRizol	ThermoFischer	Cell Lysis; DNA, RNA and Protein isolation	Separation of all the three biomolecules	No sample preparation for any omics study
Solution Based	-	Protein processing for MS	Classical method; Performed without any additional device	Time and labour intensive; No Automation; No multiomics

Table 1: A list of different bottom-up proteomics workflows with usage, advantage, and limitation

Analysis of peptides by tandem mass spectrometry

Mass spectrometry as an analytical instrument can be used to detect the mass-to-charge (m/z) ratio for ions. The obtained mass spectrums indicating the intensity of the ratio can be used to determine the elemental or isotopic signature of a sample, and masses of molecules to elucidate the chemical identity or structure of molecules. Based on the ionization source used in the mass spectrometer it can be divided into matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). Both types have found relevance in metaproteomics and the latter is widely used for shotgun proteomics. One advantage of ESI-MS is that it can be coupled to a liquid chromatography system that will enable the separation of peptides based on their hydrophobic properties using reverse-phase chromatography. Peptides are initially trapped in a trap column and then separated on a reverse-phase stationary column with a gradient of mobile phase consisting of acidified polar solvents.

In tandem mass spectrometry (MS/MS), more than one mass analysis is performed by the mass analysers upon fragmentation of the precursor ions into smaller ions to enable better identification. This is particularly useful to identify the difference

between the ions having similar m/z from the first analyser. Figure 11 shows the different steps involved in tandem MS/MS. The peptides are injected under high voltage to enable ionization by (ESI). Then the quadrupole analyser allows the selection of specific mass ions (determined by the mode of analysis and is explained later) to be further fragmented by high collision with inert gas. The fragmented ions are then analysed repeatedly. The detector records the m/z values at every stage to generate a spectrum that will be analysed at the post-analytical stage.

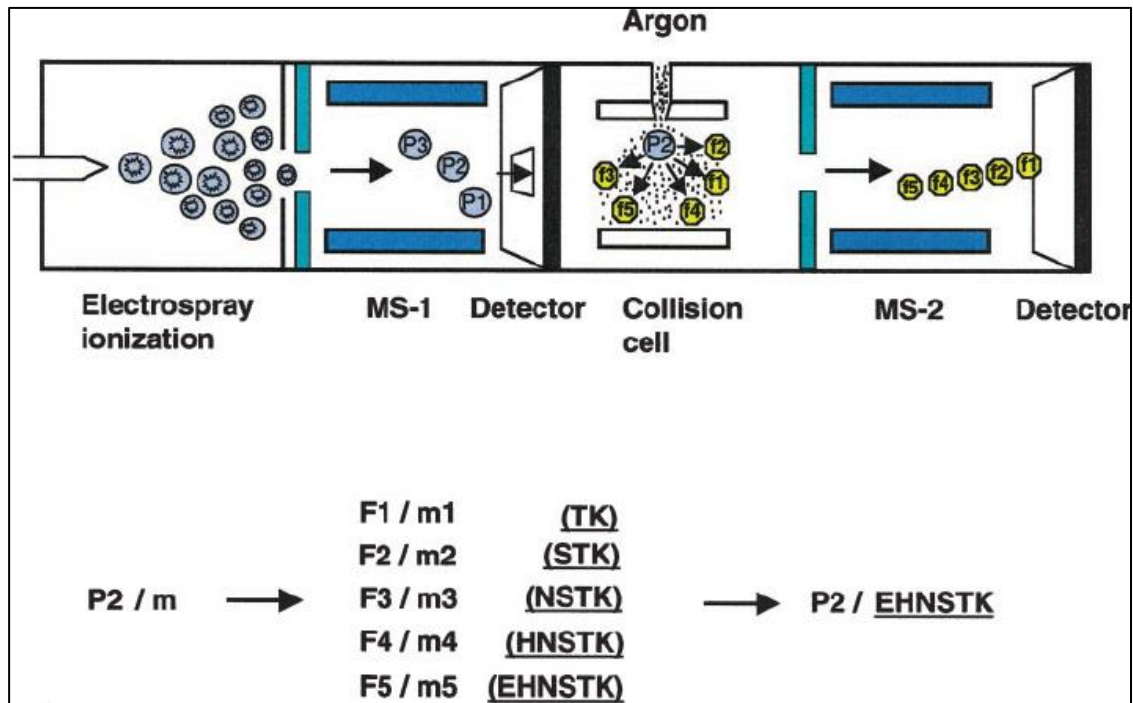


Figure 11: Schema for different steps in tandem MS/MS (Li & Assmann, 2000)

Tandem MS/MS includes the selection of the fragments that can be further analysed and the different selection procedures are known as single/multiple reaction monitoring (SRM/MRM), parallel reaction monitoring (PRM), data-dependent acquisition (DDA) and data-independent acquisition (DIA). The description of the difference between every method is shown in figure 12 and the definitions are aligned with this figure. SRM/MRM is useful to select one or more precursor ions for fragmentation and monitor/detect a selected set of fragmented ions. These assays are intended to measure a very limited number of proteins in a large set of samples making SRM an attractive technique for biomarker verification and possibly even validation (Drabovich et al., 2013). PRM allows the identification of all the fragmented ions instead of the selected one. It has a similar application in targeted proteomics and biomarker validation like SRM (Rauniyar, 2015). DDA allows the selection of multiple precursors (Top #N) based on the m/z intensity to fragmentation and monitor all the product ions. This method is the most

widely used in shotgun proteomics to identify the total proteome. DIA is like DDA, but the selection of precursors is defined through predetermined m/z windows and this can enable the selection of every precursor to be fragmented and monitored. DIA is advancing very rapidly in the field of metaproteomics and the development of data analysis software facilitates this progress. Often the DIA experiments must perform in union with DDA to obtain better results, but this trend is gradually changing. Like DDA, DIA can also be used for total proteome identification and can also enable the identification of low m/z intensity ions that are often masked in DDA experiments.

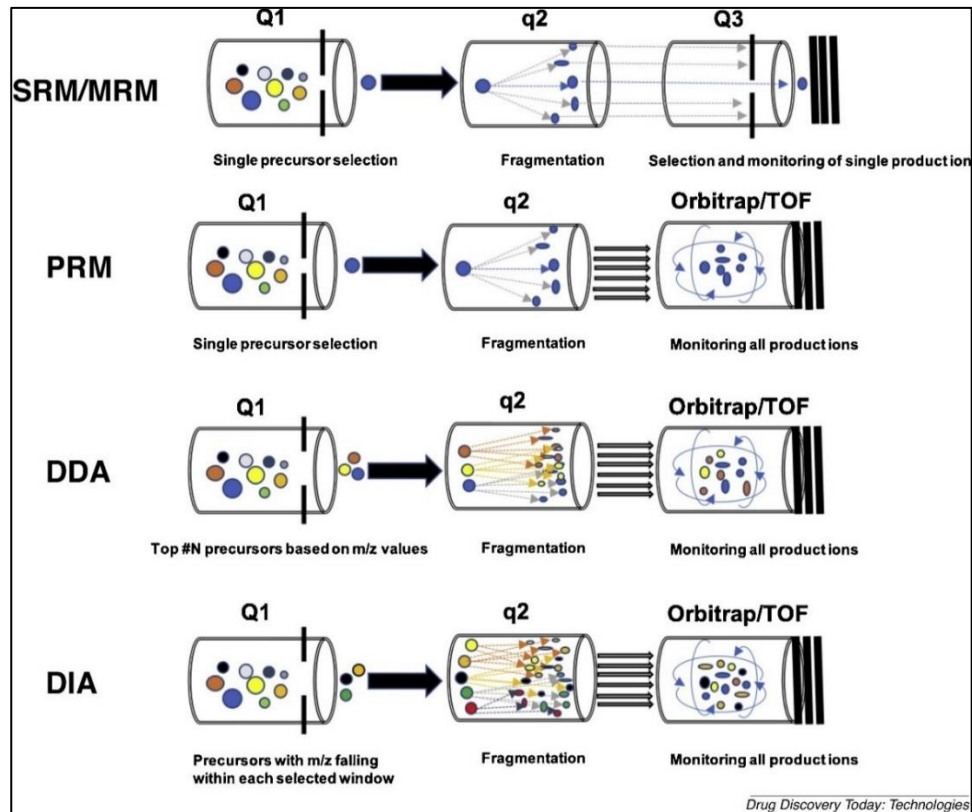


Figure 12: Difference between SRM/MRM, PRM, DDA, and DIA modes used during MS/MS (Li & Zhu, 2021)

Data analysis

Proteomic data analysis constitutes the mapping of the spectral information (m/z values) with peptide sequences and in turn aligns them to a larger protein database to comprehensively identify the protein type, associated taxonomy, and function. The first step will be to create a protein sequence database containing the amino acid sequence information for the proteome of a species. Next, these proteins are computationally digested to generate peptides m/z and their corresponding MS spectra. This can include the decoy (false-positive sequences) and target (original peptide sequences). Next, the

matches of the spectra obtained and computationally generated are performed to obtain peptide spectral matches (PSM). These scores are statistically computed to obtain the peptide sequence. The alignment of the peptides to protein is performed and scores are used to discover the matching proteins.

The key stage in metaproteomics where the species information is vaguely present is to develop a proteome database of the organisms whose proteins are being examined. Three ways that exist prominently to perform data analysis for metaproteomics are shown in figure 13. Incidentally, a protein database-free method known as the *de novo* approach can be used. The MS/MS spectra are directly read for the mass of the individual amino acids and the peptide sequence is built upon it (Hughes et al., 2010). The peptide sequences are then aligned to the protein database by using alignment tools like Basic Local Alignment Search Tool (BLAST) to identify candidate proteins. This method is utilized in software packages like PEAKS and PepNovo. Alternatively, the next two approach deals with the creation of protein databases. In the second approach, protein databases from species that resemble closely with the analysed metasample (either the composition or similar environment) can be used to create the databases. This method is not exact but is relatively faster and easier to use. The last approach is to generate expressed sequenced tags (EST) databases using the genomic or transcriptomic sequence information known as MAG of the same samples used for metaproteomics. Upon the construction of protein sequence databases, search engines that use probabilistic scoring algorithms like MASCOT or SEQUEST can be used to identify the peptide or protein sequences from MS/MS spectra.

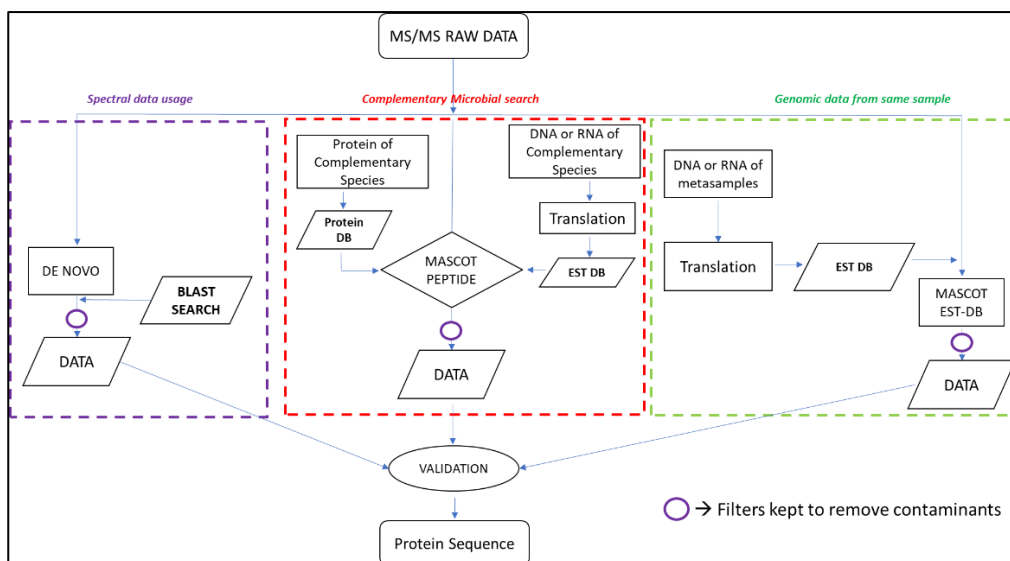


Figure 13: Different pipelines to overcome the data analysis challenges in metaproteomics

Post-processing during data analysis allows for obtaining useful conclusions relevant to metaproteomics. It necessarily involves converting the obtained protein information into useful representations with biological significance. These qualitative and quantitative results can be generated using software packages like MaxQuant (Tyanova et al., 2016), Proteome Discoverer (PD), and metaproteomics-specific packages like MetaProteomeAnalyser (Heyer et al., 2019) or MetaQuantome (Easterly et al., 2019). The different results that can be obtained are highlighted in figure 14. In differential metaproteomics, spectral counting approaches like normalized spectral abundance factor can be used for estimating abundances (Florens et al., 2006). Species abundance within a community can be estimated as a function of the relative biomass contribution for each taxon. Similarly, differential abundance estimation can be used to identify the proteins differing based on treatments, conditions, and locations. Multivariate analyses visualized by principal component analysis (PCA) plots and hierarchical clustering help characterize the differences across samples during different treatments. The functional analysis involves Clusters of Ortholog Groups which are proteins from different microbes sharing the same functional characteristics or gene ontology studies. Furthermore, studying the metabolic pathways through reconstruction can be relevant in several metaproteomic studies. Additionally, protein-protein interaction information can also be generated. These combine to also provide host-microbiome interactions. Integrating metaproteomics with other omics approaches can vastly expand the scope of the studies and is introduced in section 1.5.

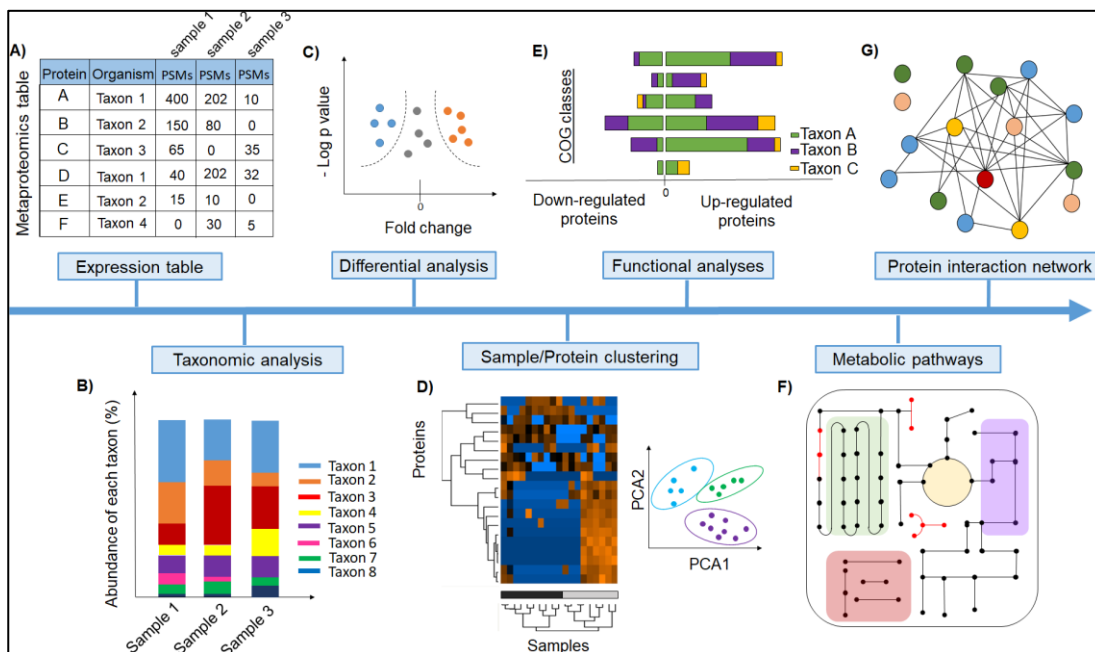


Figure 14: Results that can be obtained after metaproteomics data analysis (Salvato et al., 2021)

1.3.2. Accomplishments

Compared to metagenomics, a PubMed search for the keyword “Metaproteomics” resulted only in 987 hits from 2004 until 28-11-2022. Although proteomics methods have been existing earlier than that, the usage of these techniques for microbial community studies is not known. Metaproteomics has emerged as the technology to validate the results of metagenomics in many functional studies. Quantitative metaproteomics has become useful to determine the biomass contribution of various participating microbes within the community.

In environment microbiology, metaproteomics has contributed significantly to understanding the microbial groups from marine or freshwater, soil, wastewater, or activated sludge and acid mine drainage biofilms. Ocean water microbial communities off the coast of the Atlantic Ocean have identified more than 7600 proteins (Georges et al., 2014). Studies from freshwater for oligotrophic and eutrophic conditions have enabled the identification of 1048 proteins (Russo et al., 2016). Soil metaproteomics has resulted in a great number of publications and has gathered great attention due to its complexity in composition. Over 5000 proteins have been identified from different layers of the soil (Hultman et al., 2015). Microbes from wastewater or sludges hold key answers to the metabolic elimination of waste and studies have shown more than 5000 identified proteins (Roume et al., 2015). Acid mine drainage biofilms complicate the treatment procedure by increasing the pollutant release. Studies focus to understand the conditions that promote their growth and more than 4250 proteins have been identified (Li et al., 2014).

In human microbiome studies, a great deal of metaproteomics approaches has been applied to understand the gut, oral and vaginal microbiome. Gut microbiome studies have revealed the relationship between CRC pathogenesis and intestinal microbes. Proteins associated with iron intake or transport and oxidative stress which is indicative of cancer have been found in the study of faecal microbiomes from healthy and cancer patients (Long et al., 2020). Additionally, gut metaproteomics also revealed the microbial influence in diabetes, inflammatory bowel disorder, and obesity (Pan & Chen, 2020; Calabrese et al., 2021). Oral microbiome from saliva has been studied to identify the relationship with diseases and there have been at least 2234 proteins identified (Grassl et al., 2016). Metaproteomics studies performed on the vaginal microbiome have focussed to understand their relationship with infections (Alisoltani et al., 2020).

A key accomplishment in the field of metaproteomics is the study performed for benchmarking the steps in sample preparation, mass spectrometry, and bioinformatic analysis for a laboratory-assembled human intestinal model and a human faecal sample (Van Den Bossche et al., 2021). Results indicated variability at the peptide level due to

different sample preparation steps with less contribution from bioinformatic pipelines. The peptide level differences largely disappear at the protein level. During analysis, variability was seen for predicted community composition but similar functional profiles were obtained across workflows. This work was made possible as a result of the formation of the international community for metaproteomics (<https://metaproteomics.org/>).

1.3.3. Challenges

Due to the vastness of the proteins from intra- or inter-populations with limitations in analytical methods, metaproteomics faces multiple challenges as shown in figure 15. At the pre-analytical sample preparation stage, the key challenges are sample-specific comprehensive cell lysis and protein extraction from diverse species, efficient protocols for sub-cellular proteome fractionation (membrane, extracellular and soluble fractions), proteome fractionation based on physiochemical properties for multi-dimensional LC-MS/MS. Microbial communities can be isolated from different environments and each of them is unique with highly interfering substances (e.g. soil microbiome samples have high humic acid contents that interfere with protein purification) for a mass spectrometry study. The development of strategies to remove these interfering substances is critical for high throughput results and current methods for sample preparation needs to be improved to enable it. Protein identification from species lacking proteome databases, limited cell numbers, handling of pathogenic samples, and automation are other issues to be addressed.

At the analytical stage, the challenges include the development of *in-situ* metabolic label-free quantification methods, enhancing the separation during LC, and developing MS capable of fast scan speeds and high mass accuracies. LC and MS have been developed widely to include better analysis. Newer developments like ion mobility allow for better resolution and depth in analysis. Yet, most of these technologies are to be fully utilized for metaproteomics studies. Also, the usage of DIA-MS methods can be very attractive for metaproteomics.

At the data analysis stage, the challenges include the development of mass spectral search databases based on metagenomic and metatranscriptomics data, detection of post-translational modifications, integration of other omics approaches with metaproteomics and targeted validation of biomarkers identified in metaproteomics samples. The reliability of metaproteomics for databases to analyse the spectra forces the development of methods to process the same sample for genomics along with proteomics. Demand for functional annotation of the protein sequences is also critical. Availability of data, integration, and sharing is also key.

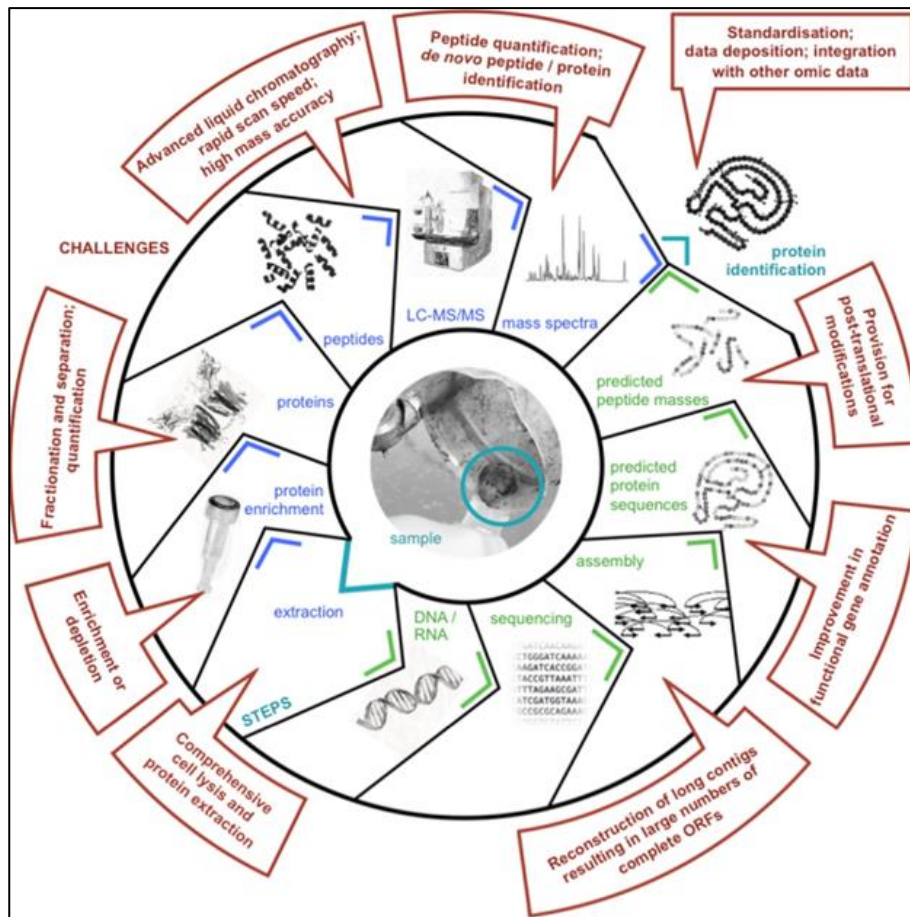


Figure 15: Challenges in different stages of the metaproteomics experiment (Wilmes et al., 2015)

Each of these challenges is the subject of intense research. In this work, we have focused on the pre-analytical steps using a microfluidic system. To better understand the developments made, we will present in the following section the specificities of the microfluidic systems.

1.4. Microfluidics technology

Microfluidic technology utilizes the precise control and manipulation of fluids that are geometrically constrained to a small space (within a sub-millimeter) at which the surface forces dominate volumetric forces. Microfluidic systems can transport, mix, separate and process fluids. They find applications in situations that require multiplexing, automation, and high throughput screening. This makes microfluidics indispensable for analytical chemistry applications. The system consists of components like micropumps and microvalves useful for supplying fluids and regulating flow direction respectively. Typically, micro means at least one of the following features:

- Small volumes (μl , nl, pl, fl)
- Small size
- Low energy consumption
- Microdomain effects

1.4.1. Different fabrication processes and microfluidic systems

Mould-based microfluidic fabrication methods are very popular as they offer the advantage to utilize polymers compatible with biological applications, offers lower production cost for mass manufacture than photolithography, and are easier to replicate. The resolution depends on the mask used and can reach up to 6nm (Waldner, 2010). After the creation of the moulds, polymers are used to create chips with microchannels. The most commonly used polymer chemistries include vinyl (polyethylene) and acrylates (poly (methyl methacrylate)), epoxy resins (SU-8), thiol-enes (Norland optical adhesive), polyurethanes, and siloxanes (polydimethylsiloxane (PDMS)). The basic steps in this process from mould fabrication to device fabrication are outlined in figure 16. The same workflow will be used for ChipFilter fabrication explained in this work (see section 2.1.2.).

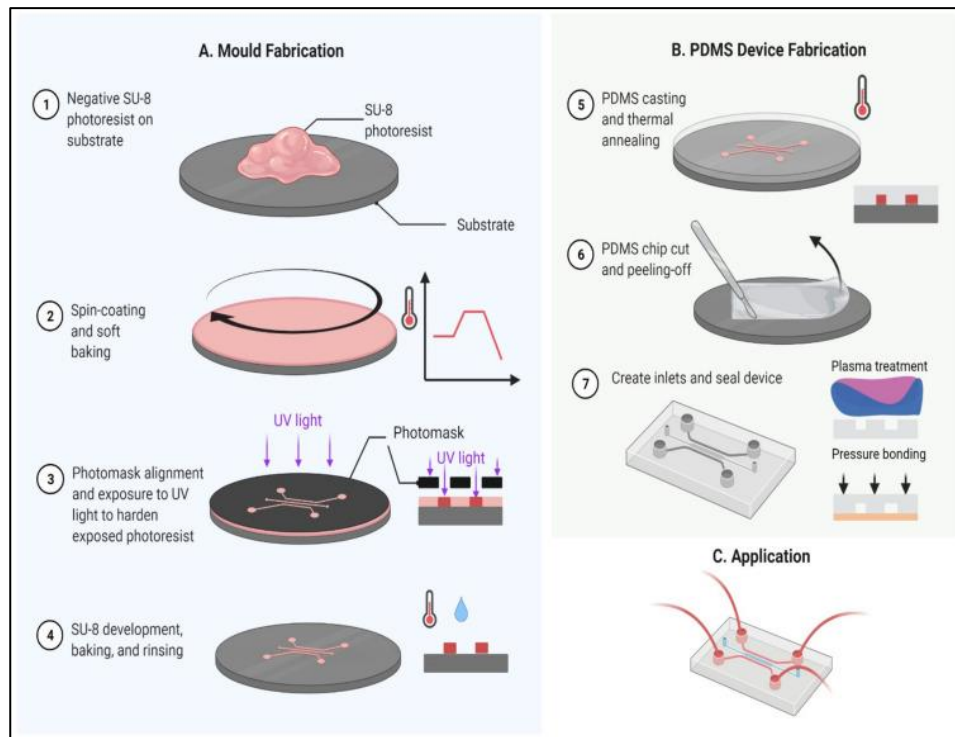


Figure 16: Representation of the different steps involved in the mould fabrication and PDMS device fabrication (Scott & Ali, 2021)

Different kinds of microfluidic systems have been developed and made useful for biological applications. The different types of systems are presented below with their advantages and uses.

Open microfluidics

The fluids inside these systems are exposed to the environment since at least one of the boundaries of the microchannels is open and thereby maintains a constant air-to-liquid interface. The open boundary is constantly in contact with air or other liquid (Berthier et al., 2016; Pfohl et al., 2003; Kaigala et al., 2012). Advantages include accessibility to the flowing liquid, larger liquid-gas surface area, minimized bubbling, and operation without external pumping systems (Berthier et al., 2016; Kaigala et al., 2012; Lade et al., 2018; Li et al., 2017 (A); Casavant et al., 2013). The fabrication of open systems is easier than closed systems as they do not require an additional step for bonding. All the regular methods like soft-lithography, milling, thermoforming, and hot embossing can be used for fabrication. Additionally, surface modifications by UV treatment or deposition can be done easily on the channels as their boundaries are open (Guckenberger et al., 2015; Truckenmüller et al., 2002; Jeon et al., 2011). Disadvantages include contamination, limited flow rate, and evaporation (Kachel et al., 2014).

Continuous-flow microfluidics

In continuous-flow microfluidics, the fluid flow is controlled through the channels or porous media without breaking the fluid flow. The fluid flow is controlled by external pressure sources, pumps (external or integrated), or electrokinetic elements (Chang & Yeo, 2009). These systems offer advantages like easy implementation and lower protein fouling problems. They find applications in chemical separation and catalysis. Disadvantages include difficulty in integration, closed channels, and permanently etched microstructures. The ChipFilter used in the present study is a continuous-flow microfluidics system.

Droplet-based microfluidics

The generation and manipulation of discrete droplets inside microchannels using two immiscible liquids interfaces are achieved in droplet-based microfluidics. Microdroplets developed in these systems allow for handling small volumes of fluids that can be useful for encapsulation, sorting, mixing, and sensing (Chokkalingam et al., 2013). Advantages include compartmentalized reactions inside the droplet to prevent sample diffusion or contamination, handling femtolitre to nanolitre volumes, monodispersity of the droplets allowing identical conditions, enhanced mixing, and a high-throughput process.

Digital microfluidics

Digital microfluidics utilizes the electrowetting properties of the droplet to manipulate their dispersion, movement, mixing, and storage on a platform insulated with electrodes. Altering the voltage will allow the movement of the droplets. The three most commonly used principles used to generate and manipulate microdroplets in a digital microfluidic device are electrowetting, dielectrophoresis, and immiscible-fluid flows. It offers a similar advantage as droplet microfluidics with additional capacity to control every droplet generated in the channel. Several lab-on-chips use digital microfluidics (Wang et al., 2017). Recent advances in digital microfluidics have shown droplet manipulation using magnetic force (Zhang & Nguyen, 2017), surface acoustic waves (Shilton, 2014), optoelectrowetting, and piezoelectricity (Shemesh et al., 2010).

Paper-based microfluidics

Paper microfluidics consist of hydrophilic cellulose or nitrocellulose membrane that acts as capillaries to transfer liquids from the inlet through to the desired target. The advantages are portability, cheapness, and user-friendly in medical diagnostic systems (Alsaeed & Mansour, 2020).

Microfluidic properties

In microfluidics, several physical phenomena govern the behaviour of the particles and fluids in the system. The effects that become dominant in microfluidics include Reynold's number, laminar flow, diffusion, fluidic resistance, surface area to volume ratio, and surface tension. The fluid properties mentioned below make the technology advantageous to biological applications.

Reynold's Number

Reynold's number (Re) of a fluid flow is used to describe the flow regime if it's either laminar (Re < 2300) or turbulent (Re ≥ 2300). The Reynolds number can be calculated by

$$\text{Re} = \frac{\rho v D_h}{\mu}$$

where ρ is the fluid density, v is the characteristic velocity of the fluid, μ is the fluid viscosity, and D_h is the hydraulic diameter that is dependent on the cross-sectional geometry of the channels.

Laminar Flow

Laminar flow allows the particle in a fluid flow with a certain velocity to have a non-random function of time. In microchannels, due to their small size, the flow is almost always laminar (White, 1991). This phenomenon (except by diffusion) helps to avoid the mixing of two or more streams flowing in contact with each other and allows the creation of fluid packets that are stable to be transported in a controlled manner in the stream.

Diffusion

The movement of concentrated particles in a volume by Brownian movement to spread out over time to the entire volume of the solution to make it constantly distributed is called diffusion. In one dimension, it can be represented by the equation,

$$D = \frac{d^2}{2t}$$

where d is the distance, a particle moves in a time t , and D is the diffusion coefficient of the particle. Since the distance in microchannels is small the time required for diffusion to be achieved is also short. Microchannels enabling different diffusion times for the fluids can be utilized for the creation of complex concentration gradients (Dertinger et al., 2001; Jeon et al., 2000).

Fluidic resistance

Fluid resistance or the drag in a microchannel is the resistance created by the fluid and is given by the equation,

$$Q = \frac{1P}{R}$$

where Q is the flow rate, $1P$ is the pressure drop across the channel, and R is the channel resistance. The geometry of the channels determines the outcome of the resistance.

Surface area to volume ratio

The surface-to-volume ratio in microchannels is very high and this factor can be very much more useful than the conventional macroscale operation. This makes capillary electrophoresis more efficient in microchannels due to better heat removal operation. At the same time, a large surface-to-volume ratio will increase the rate of diffusion and can lead to adsorption to channel surfaces (Manz et al., 1994; Locascio et al., 1999)

Surface tension

Surface tension forces at the microfluidic level are significant to propel the flow of the liquid and the effects of surface tension begin to overcome Newtonian forces such as gravity and inertia. It is a result of the cohesion energy between liquid molecules at the liquid/gas interface. Surface treatments of the microchannels can enable lower tension and hence altered wetting properties.

1.4.2. Microfluidics for microorganisms

Microfluidic technology has found great use in microbiology and addresses its challenges. In this part, some of the applications of microfluidics in microbial culture, sorting, pathogen detection, and antibiotic screening/resistance are discussed

For microbial culture

The growth of cells in microfluidic offers several advantages over the conventional flask culture with less culture volume, the precise selection of the microenvironment (close to natural), single-cell culture, and biofilm formation studies (Nilsson et al., 2009; Rusconi et al., 2014; Yawata et al., 2016). The cells can be grown in different dimensions 1D, 2D (single-cell level), and 3D (population level) using microfluidic devices as shown in figure 17. Several microfluidic systems droplets (Park et al., 2011), microwells, membrane separated (porous for cellular metabolites but not to cells) (Burmeister et al., 2019), or hydrogel encapsulated has been developed (Moffitt et al., 2012). Cells cultured under these confinements are used to study cell-cell interactions like quorum sensing, cross-feeding dynamics, and habitat competition. Further, bacterial and eukaryotic cell

interactions can also be studied for example interaction of lung epithelial cells and bacteria (Li et al., 2016).

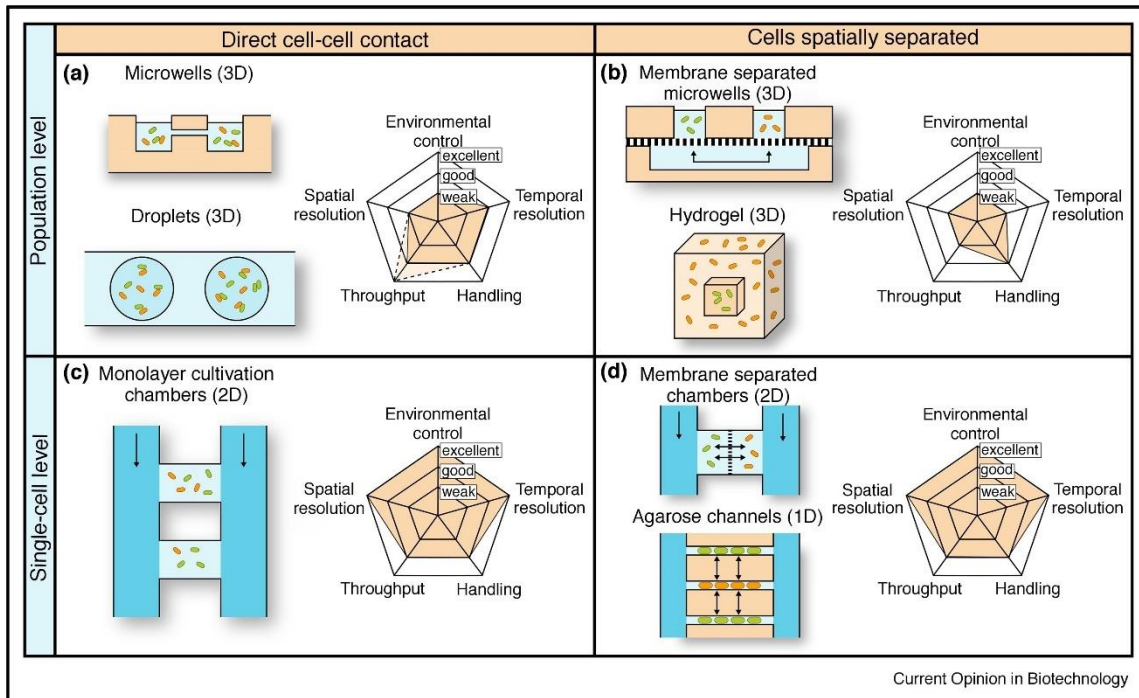


Figure 17: Different microfluidics systems for microbial culture and interaction studies (Burmeister & Grünberger, 2020)

For cell sorting

Droplet microfluidics has been greatly utilized for the function of sorting microbial cells from a population to generate single cells that can be used for screening of libraries as shown in figure 18, cultured in droplets to study evolution (Rakszewska et al., 2014), or used for single-cell analysis. The sorting methods are based on labelled detections or label-free detection methods. Labelled detection methods include fluorescence-activated droplet sorting, and absorbance-activated droplet sorting, whereas label-free detection methods include electrochemical-based droplet sorting, mass-activated droplet sorting, Raman-activated droplet sorting, and nuclear magnetic resonance-based droplet sorting. Droplet-based microfluidics paves a cheap and convenient way for ultra-high throughput *de novo* synthetic enzyme screening.

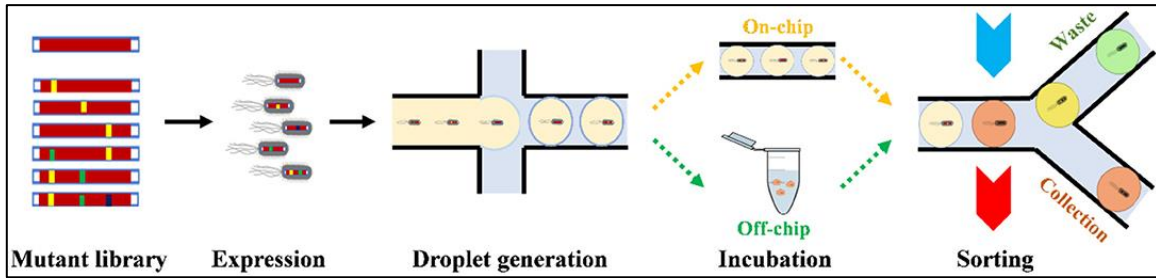


Figure 18: Application of droplet microfluidics systems for microbial sorting and mutant selection studies (Fu et al., 2021)

For pathogen detection

Microfluidics-based sensors and detection systems for pathogens and cellular components are attractive applications. These systems are capacitated to perform lab-on-chip operations with low sample volume at high speed in a high-throughput manner, thus making them ideal for point-of-care applications. A complex shutter flow device containing micropump and microvalves was designed with 48 units to perform screening of dengue virus with a detection limit of 100 pM was achieved in only 90 s using 1 μ l of the sample (Huang et al., 2010). Microfluidic devices coated channels with active molecules such as antibodies, antimicrobial peptides, lectins, aptamers, bacteriophages or molecularly imprinted polymers can also help to detect foodborne pathogens (figure 19). Droplet, digital, lab-on-disk, and paper-based microfluidic devices have been used in many clinical pathogen detection assays.

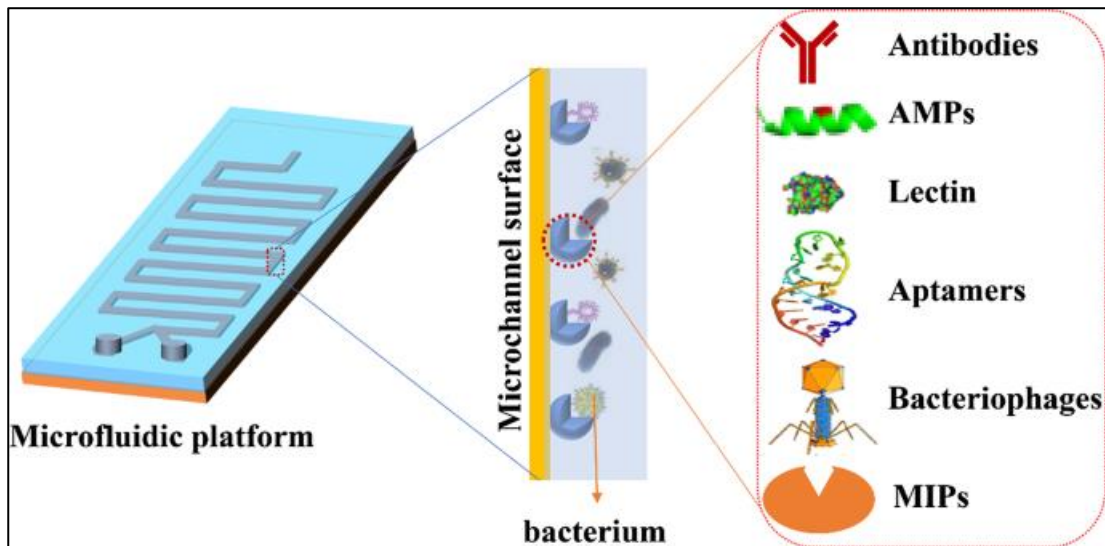


Figure 19: Application of microfluidic channels for microbial detection through surface coating with antibodies, antimicrobial peptides (AMPs), aptamers, carbohydrate-binding molecules (lectin), bacteriophages, and molecularly imprinted polymers (MIPs) (Mi et al., 2022)

For antibiotic screening and resistance

The development of microfluidic systems has led to significant insights into the single-cell dynamics of microbes, performing antibiotic susceptibility testing, isolating targeted microbes by cultivating on-chip and exploring new antibiotics (Liu et al., 2017). High throughput screening of antibacterial drugs and a broad range of on-chip approaches for antibacterial resistance detection (Qin et al., 2021).

1.4.3. Microfluidics for sample collection and processing

In microbiome research, challenges arise at different stages during sample processing which includes cell collection from highly dispersed environments like aerosol, cell lysis, single-cell sample processing, and automation. As the sample collection and processing stage is crucial to preserving the biomass and subsequent analysis, several microfluidic-based technologies have been developed for it.

Airborne microbial collection

Aerosols contain many bacterial, fungal, and viral particles that are associated with diseases. Detection of these pathogens is complicated by the fact that they can be highly dispersed and require an intensive collection process. Unfortunately, there are not many standard methods for processing these samples and hence screening can be complicated. Microfluidics has allowed the development of devices to aid in sample collection and simultaneous processing for the screening of pathogens. Microfluidics used for collection can be structurally classified into herringbone-based microfluidic chips (Jing et al., 2013; Jing et al., 2014; Li et al., 2018; Bian et al., 2016), centrifugation-based microfluidic chips (Ma et al., 2016; Hong et al., 2015; Choi et al., 2017), droplet-based microfluidic chips (Damit, 2016) and miniaturization of traditional collection procedures like electrostatic precipitation (Ma et al., 2016) and filtration (Liu et al., 2018 (A)) (figure 20).

The design used for the airborne microbial collection is diverse. In the case of the herringbone structure, the bioaerosols were continuously sucked within the structures using a pump and there is a vortex created in the laminar-flow bioaerosols allowing the particles to be captured onto the microchannels. The centrifugation-based devices rely on the inertial differences to separate and collect the different-sized particles in bioaerosols. Droplet-based chips can be used to encapsulate the bioaerosols and thereby trap the particles within them. In the collection procedure using electrostatic precipitation, the charged bioaerosols were captured in a half-open microchannel by an electrostatic field and resuspended in the collection liquid. In filtration-based methods, microfilters were embedded in PDMS to trap the particles in the bioaerosols.

Furthermore, the detection of these species was enabled by nucleic acid-based methods like polymerase chain reaction (Jiang et al., 2014), isothermal and chimeric primer-initiated amplification of nucleic acids (Inami et al., 2009), or loop-mediated isothermal amplification (Jiang et al., 2016; Liu et al., 2018 (B)), and protein-based Immunoagglutination assay (Kwon et al., 2014) or immunoassay based on field effect transistor (Shen et al., 2011), chemiluminescence (Fountain et al., 2014; Coudron et al., 2019) or fluorescence (Jing et al., 2014; Li et al., 2018). These devices allowed the collection and detection of microbial cells like *Escherichia coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Vibrio parahemolyticus*, *Staphylococcus epidermidis*, *Pseudomonas aeruginosa*, H1N1/2009 Virus, Influenza H3N2 viruses, and *Aspergillus niger*.

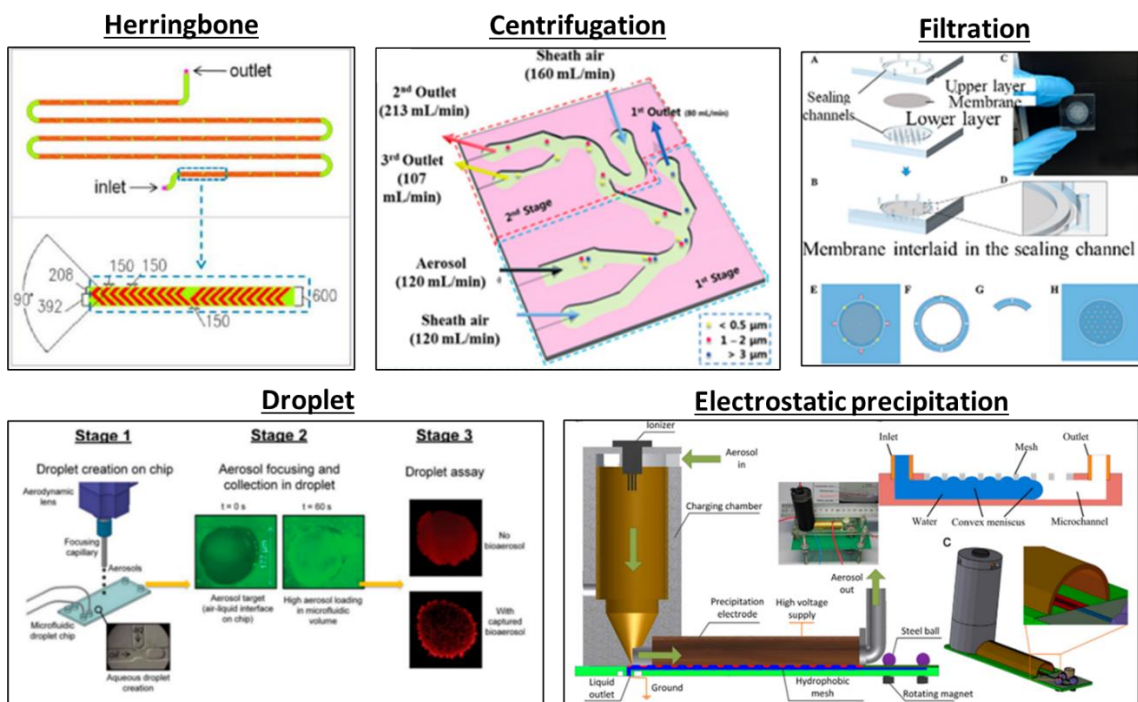


Figure 20: Application of different microfluidic devices for microbial collection from bioaerosols. (Jing et al., 2013 ; Hong et al., 2015 ; Bian et al., 2016 ; Damit et al., 2016 ; Ma et al., 2016)

Sample preparation

The most important sample preparation step is the cellular lysis to release different biomolecules. In microbial samples, the complexity of the cell lysis process is increased due to the composition of the cell wall/membrane of the microbes. Cell lysis can be achieved in microfluidic devices by chemical, mechanical, electrical, thermal, laser, acoustic, and electrochemical methods. The performance of each method is mentioned in table 2.

Chemical lysis utilizes alkaline reagents or detergents for cell lysis. Detergents commonly used include ionic detergents like sodium dodecyl sulfate (SDS) or non-ionic detergent like Triton-X. In a study to compare the efficiency of continuous flow lysis, conditions like detergents composition, flow rate, contact times, and cell-to-lysis solution volume ratios indicated that commercially available non-ionic surfactant (BPER) with lysozyme proved to be very effective (Grigorov et al., 2021). In the present study, Chemical lysis is preferred for the simple and effective operation of a continuous-flow microfluidic system. Further, the non-ionic detergent Octyl b- D -glucopyranoside (ODG) is used. It is effective in the solubilisation of integral membrane proteins.

Mechanical methods involve the usage of sheer force to cause cell lysis. The most common way in microfluidic devices is to include nanoscale obstacles in the microchannels. This can be done by using sharp nanostructures (Carlo et al., 2003), porous silica monoliths causing disturbances in the flow (Han et al., 2019), constriction in the channels (Huang et al., 2019), or introducing compressive stress through polymer deflection between microchannels (Kim et al., 2007).

Electrical methods or electroporation utilize an electrical field above the threshold of transmembrane potential to create pores on the membrane of the cells and thereby releasing intercellular contents. The electric potential is maintained through gold or platinum electrodes. Alternating current is preferred over direct current to avoid bubbling effects (Mernier et al., 2010). Shape-selected cell lysis under electric fields has also been developed allowing preferential cell lysis from a mixture (Kremer et al., 2015).

Laser lysis utilizes the fluid motion created by a focused laser to break the cell membrane. A laser pulse focused at the buffer interface of a cell solution produces a localized cavitation bubble. The expansion and the subsequent collapse of the bubble induce fluid dynamic forces that lead to the breakdown of the cell membrane (Nan et al., 2014).

Thermal lysis uses high temperatures to denature the membrane proteins, thereby denaturing the cells to access intracellular components. Temperature sensors are integrated to control the temperature and heating is performed through ohmic heating as it consumes less power and is easily miniaturized for microfluidics chips. A key advantage of these devices is that temperature control can also allow for the enzymatic amplification of the nucleic acids obtained after cell lysis (Lee et al., 2005).

Finally, the use of acoustic waves generated through piezoelectric crystal substrate is effective in the lysis of *E. coli* cells (Lu et al., 2019). In electrochemical lysis, saline solution is used to generate hydroxide ions (OH-) at the cathode chamber that is

separated by a negatively charged ion exchangeable polymer diaphragm to maintain the high pH level for efficient cell lysis in the cathode chamber (Lee et al., 2010).

Lysis Type	Efficiency	Lysis Time	Technical Difficulty	Cost
Chemical	High	Slow/Moderate	Low	Low
Mechanical	Medium	Moderate	Medium	Medium
Electrical methods	High	Fast	High	High
Thermal	Medium	Moderate	Medium	Medium
Laser	High	Very Fast	Very High	Very High
Acoustic	High	Very Fast	High	High
Electrochemical	Medium	Moderate	High	Moderate

Table 2: Comparison of different microfluidic cell lysis methods on efficiency, lysis time, technical difficulty, and cost. (Grigorov et al., 2021)

Subsequent steps in sample preparation are dedicated to the preparation and analysis of certain biomolecules like nucleic acids, proteins, or metabolites. Dedicated workflows using microfluidics systems exist and some of the most relevant methods and devices are discussed under the next title.

1.4.4. Microfluidics for genomics and proteomics

Microfluidics has been utilized in multiple ways for the detection of biomolecules from microbial cells. A key detection strategy of the complete genome or proteome is through their respective sequencing. Microfluidic technology has been either utilized for entire omics approaches or partially at specific places to enhance the workflow. Some of the relevant technologies are introduced below.

Genomics

The separation of nucleic acids (DNA or RNA) from other components can be achieved through various microfluidic devices that can be broadly classified as (i) silica-based techniques, (ii) paramagnetic beads-based techniques, (iii) surface modification techniques and (iv) liquid-phase isolation techniques (Reinholt & Baeumner, 2014).

Silica-based techniques benefit from the non-specific binding of nucleic acids with silica and are enhanced by the usage of chaotropic agents to extract nucleic acids. Devices with packed silica beads (Easley et al., 2006), sol-gels incorporating silica beads (Legendre et al., 2006), porous polymer monoliths incorporating silica beads (Mahalanabis et al., 2009), fabricated silica microstructures (Cady et al., 2005), silicate-based porous monolith (Kashkary et al., 2012), and silica membrane (Chen et al., 2010) have been developed.

Paramagnetic beads-based techniques use surface-coated paramagnetic beads to bind nucleic acids and then be enriched along with the bound nucleic acids using an applied magnetic field. Devices with silica-coated paramagnetic beads (Duarte et al.,

2011), paramagnetic beads with switchable charges (Liu et al., 2013), paramagnetic beads coated with oligo-dT (Berry et al., 2011), and paramagnetic beads coated with specific sequences (Wang et al., 2011 (A)) have been developed.

Surface modification can be done using oligonucleotides or other surface-modifying agents that will allow for the non-specific binding of nucleic acids under certain buffer conditions. Devices with oligonucleotides on polymer surfaces (Reinholt et al., 2014), chitosan-coated beads (Hagan et al., 2011), aluminium oxide-coated membranes (Kim et al., 2010 (A)), photoactivated poly-carbonate surfaces (Park et al., 2008), and amine-coated membranes (Nakagwa et al., 2005) have been developed.

Liquid-phase isolation techniques utilize mobility in an electric field and solution chemistry to separate nucleic acids from the sample contaminants. Devices with electrophoretic techniques (Vulto et al., 2010) and isolation with an organic liquid (Zhang et al., 2013) have been developed.

Microfluidic devices developed for genomics sample preparation applications can be grouped into devices for (i) single-cell or low-concentration samples that require whole genome sequencing, (ii) targeted sequencing, and (iii) routine library preparation step miniaturization. Devices have been developed for performing other closely related studies like digital PCR (Ottesen et al., 2006), transcriptomics (Streets et al., 2014), and chromatin immunoprecipitation (Oh et al., 2009; Wu et al., 2009).

Whole genome sequencing by using microfluidic technology can be achieved through multiple displacement amplification, PicoPLEX in a droplet, and multiple annealing and looping-based amplification cycles (MALBAC). Multiple displacement amplification is a non-PCR-based isothermal amplification that uses exonuclease-resistant random primers and strand displacing ϕ 29 polymerase to amplify femtograms or picograms of DNA templates with lengths greater than 10 kb (Dean et al., 2002). There are four types of microfluidic multiple displacement amplification methods micro-chamber (Marcy et al., 2007), droplet (Fu et al., 2015), micro-well (Gole et al., 2013), and gel (Xu et al., 2016). Multi-step PicoPLEX whole genome sequencing protocol was merged into multiple droplets based microfluidic device to analyse single microbes for cell sorting, cultivation, quantitative PCR, and whole genome sequencing (Leung et al., 2012). Microfluidic devices utilizing MALBAC technology have been developed (Yu et al., 2014). MALBAC technology utilizes primers capable of allowing the amplicons to have complementary ends and thereby allowing them to loop and prevent DNA from being copied exponentially. This will generate quasilinear amplification.

The targeted sequencing approach involves the enrichment of target sequences in the genome before sequencing. The three widely used methods for sequencing are

based on PCR, molecular inversion probes, and hybrid capture. Droplet microfluidics has been used to perform multiplexed PCR for target enrichment (Tewhey et al., 2009). Molecular inversion probes allow the circularization and amplification of the target sequences and it is been performed with microfluidic devices (Carrascosa et al., 2014). Hybrid capture utilizes probe-based hybridization of target sequences and based on the reaction phase it can be divided into on-array capture and in-solution capture. Commercial products from Roche and Agilent utilize this technology in their microarray systems.

In library preparation for NGS, key steps like DNA fragmentation, ligation, integrated library preparation, and quality control can be performed using microfluidic technology. DNA fragmentation in chips can be performed by sonication (Tseng et al., 2012) or enzymatic treatment using DNase (Anderson et al., 2000). The attachment of DNA fragments to the flow cell is enabled by ligase activity. Ligation-enabling chips are developed by utilizing flow channels for enzyme, vector, and DNA fragments (Wook Hong et al., 2006) or by electrowetting-on-dielectric chips (Lin et al., 2010). Automation of library preparation by removing enzymes, buffers, and small molecules was achieved in microfluidic chips using digital microfluidics systems integrated with AMPure XP magnetic beads (Kim et al., 2011). Library fragment size was determined using a droplet-based digital microfluidic system, capillary-based reagent delivery unit, and quantitative capillary electrophoresis module (Thaitrong et al., 2012).

Proteomics

Microfluidic technology has been used to develop chips that can perform preconcentration of proteins or peptides to complete preanalytical sample preparation steps. In the area of sample preparation and preconcentration, microfluidics technology based on microbeads, monoliths, and membranes has been widely used.

Micro or nano-sized particles are packed within microfluidic channels as seen in conventional chromatographic or solid phase extraction columns. These beads deposited in the microchannels may require the creation of special geometries to keep them in place. The advantage of the beads is that they can be chemically modified on their surface to increase the scope of application. In one application, integrated microfluidic chips consisting of beads coated with trypsin and immobilized metal affinity chromatography beads were packed into a microchannel to perform proteolysis and phosphopeptide enrichment (Yue et al., 2006). Additionally, magnetic beads were used in digital microfluidic systems to miniaturize conventional SP3 sample preparation. Complete sample preparation for Jurkat T cells led to the identification of 2500 proteins from 500 cells and about half the number of proteins from just 100 cells (Leipert & Tholey, 2019).

Monoliths consist of a single rod with a porous structure made of mesopores (2-50 nm) and microporous (>50 nm). They are useful as flow channels for solvents and provide high surface area for analytes. Furthermore, the chemical nature of the monolith can be changed to maintain charge states or hydrophobic/hydrophilic balance (Dziomba et al., 2017). The monoliths can be synthesized using polymers, silica, and organic-inorganic hybrid. Octadecyl monolithic silica microchip was used to preconcentrate milk proteins (Alzahrani & Welham, 2013).

Microfluidics devices have been made to perform separation as seen in liquid chromatography, electromigration (capillary and microchip electrophoresis), and field-flow fractionation methods. The miniaturization of LC is very attractive as it has advantages like smaller sample and reagent volumes, fast and inexpensive processing, compounds without dead volume, and the possibility of multiplex analysis. Current developments use electroosmosis and pressure to drive the analytes in a microfluidic chip with continuous beds (Ericson et al., 2000). A chip-based two-dimensional LC was developed by merging injection, separation, and detection features on a fused silica chip in a dead volume-free manner (Piendl et al., 2020). Capillary electrophoresis and microchip electrophoresis are seen as counterparts to the LC-based systems as their separation have several advantages. Chip-based capillary zone electrophoresis system was developed to enable charge-based separation of peptide mixtures and analyse by MS (Dykstra et al., 2021). Other capillary electrophoresis methods like isoelectric focusing, affinity electrophoresis, electrokinetic chromatography, and electrochromatography have been used in microfluidic chips for protein/peptide separation. Separation in field flow fractionation occurs without a stationary phase under an applied field like hydraulic, thermal, electric, or magnetic perpendicular to the direction of transport of the sample which is pumped through a long and narrow channel (Plavchak et al., 2021).

Membranes are the most commonly used solid support for analyte preconcentration in microfluidics. It can be applied for the preconcentration of small molecules to biopolymers. Thus, it can suitably be called a “semi-permeable barrier”. Membranes used in the devices can be neutral or charged (Kim et al., 2010 (B)). Neutral membranes have molecular weight cut-off that restricts the passage of particles greater than the cut-off. In charged membranes, the separation is accomplished by the concentration polarization effect. Due to this effect, a concentration gradient at the membrane/solution interface is created leading to the selective transfer of some species through the membrane under an applied voltage.

The microfluidics device utilized in this work utilizes a neutral membrane with a molecular weight cut-off of 10kDa allowing for the formation of a reaction chamber with 0.6 μl volume. The design is shown in figure 21. This device was demonstrated to perform

the sample preparation steps (protein separation, reduction, alkylation, washing, and proteolysis) to facilitate direct coupling to LC-MS/MS for analysis. Over 4000 proteins were identified from 28, 000 HT29 cells (Ndiaye et al., 2020). This device was originally restricted to eukaryotic cells and lysates, but the present work will adapt this technology for metaproteomics sample preparation along with nucleic acid retrieval. Fabrication and workflow for operation are well explained in chapter 2.

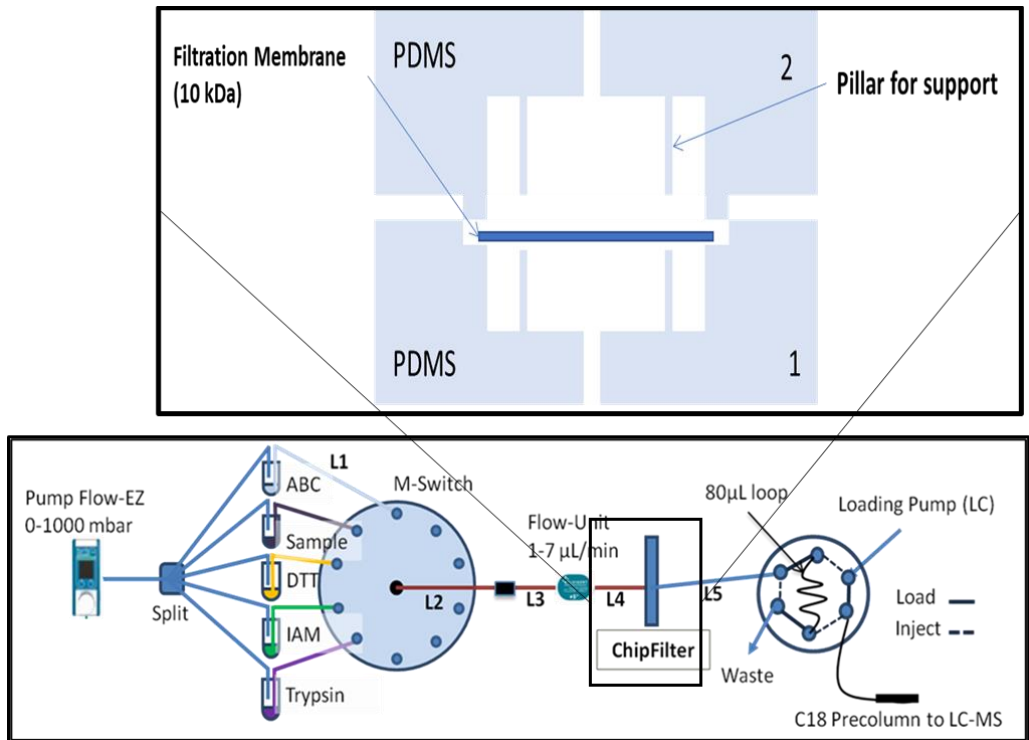


Figure 21: Microfluidic device (ChipFilter) design for bottom-up proteomics sample preparation. (Ndiaye et al., 2020)

1.5. Multiomics approaches in microbiome studies

Multiomics or integrative omics or pan omics is a strategy to integrate more than one omics study that includes genomics, transcriptomics, proteomics, metabolomics, and microbiomes. The integration of multiple datasets for the same sample allows mapping the state of the organism at different layers that will answer questions on origin, function, regulation, and interaction. This supreme knowledge about the global activity of the cell under a certain environment allows for predicting their behaviours. Systems biology allows the integration of different data sets to mine for useful inferences. The number of publications registered in PubMed under “multi-omics” has increased from 23 in 2012 to 2731 in 2022. This magnitude of growth in research on the topic demonstrates its relevance of this topic.

Single-cell multi-omics is an important field of study as it can reduce the complexity that generally arises from population studies, but at the same time, the efficiency in sample preparation for a few numbers of cells to perform different omics studies is challenging (Hu et al., 2018). Methods for single-cell studies are being extensively studied as shown in figure 22 and microfluidics will hold a key place in these studies as it can efficiently process low sample amounts. Current multi-omics studies in single-cell are focussed on linking genomics (scDNA-seq) and transcriptomics (scRNA-seq) variation using gDNA–mRNA sequencing, multi-omics analysis in cancer, linking epigenetic and transcriptomic variation and relating expression of RNA and proteins (Macaulay et al., 2017).

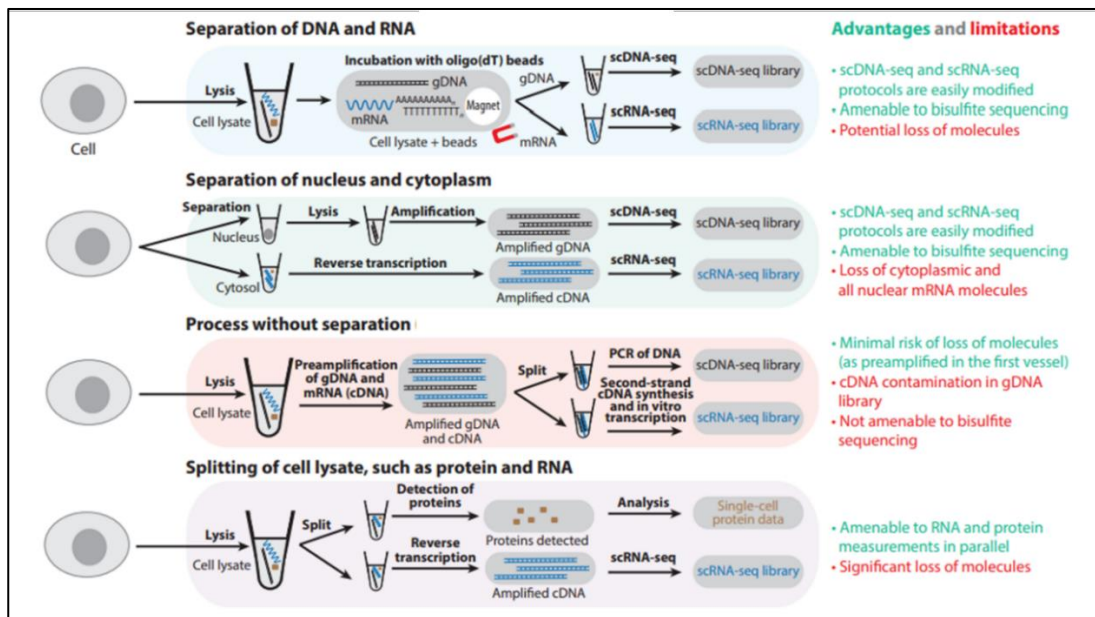


Figure 22: Methods for the processing of single cells for multi-omics studies (Chappell et al., 2018)

Successful sample preparation, multi-omics analysis, and data integration for tissues or *in-vitro* cell lines have been done (Gutierrez et al., 2018). These studies have led deeper understanding of disease pathogenesis and led to the development of precision medicine (Pammi et al., 2022). Advances in immune-mediated diseases (Ota & Fujio, 2021), cancer (Subramanian et al., 2020), and metabolic diseases (Hu & Jia, 2021) have been contributed by multi-omics. In immune-mediated diseases like rheumatoid arthritis or coronary artery disease, integration of clinical, genomic, and transcriptomics data has led to the prediction of treatment response and identify biomarkers. In the case of a rare paediatric cancer known as retinoblastoma, a system biology approach integrating transcriptomics and metabolomics has revealed cancer metabolic signatures and metabolic targets (Sahoo et al., 2019). A similar study integrating transcriptomics, proteomics, and metabolomics in RB-loss (tumour suppressor) has shown metabolic functions of the cancerous cells (Rajasekaran et al., 2021). Metabolic diseases including type 2 diabetes mellitus, non-alcoholic fatty liver disease, and metabolic syndrome have been studied by multi-omics approaches. Data sets for whole genome sequences, clinical tests, metabolomes, proteomes and microbiomes sequencing collected from 108 patients contributed greatly to developing personalized dietary interventions for metabolic diseases (Price et al., 2017).

Microbiome studies are the greatest benefactors of multi-omics. Unlike the other studies that use model organisms for understanding dysbiosis, microbiome deals with non-model organisms and their function under a certain environment. The involvement of more than one omics study is necessary to identify the microbial type (metagenomic) and associate their functions (metatranscriptomics, metaproteomic, metabolomics) to describe the phenomena of a microbial community (figure 23). Most of the highlighted works from earlier sections have utilized multi-omics approaches to identify and highlight metabolic pathways functional in a community. Another interesting work performed on the 880 microbial community samples collected for the Earth Microbiome Project was subjected to the amplicon (16S, 18S, ITS) and shotgun metagenomic sequencing with untargeted metabolomics sequencing (liquid chromatography-tandem mass spectrometry and gas chromatography-mass spectrometry). The study provided reference databases and the diversity in metabolic pathways (Shaffer et al., 2022). Multiomics performed in humans has been very useful to understand host-microbe interactions. Numerous studies have been published on the gut microbiome with disease and well-being. Multiomics has also helped to understand the association between the gut microbiome and ocular pathogenesis like autoimmune uveitis, age-related macular degeneration, and glaucoma (Xue et al., 2021). Multi-omics approaches in microbiome studies are inevitable due to the merits it brings to the study.

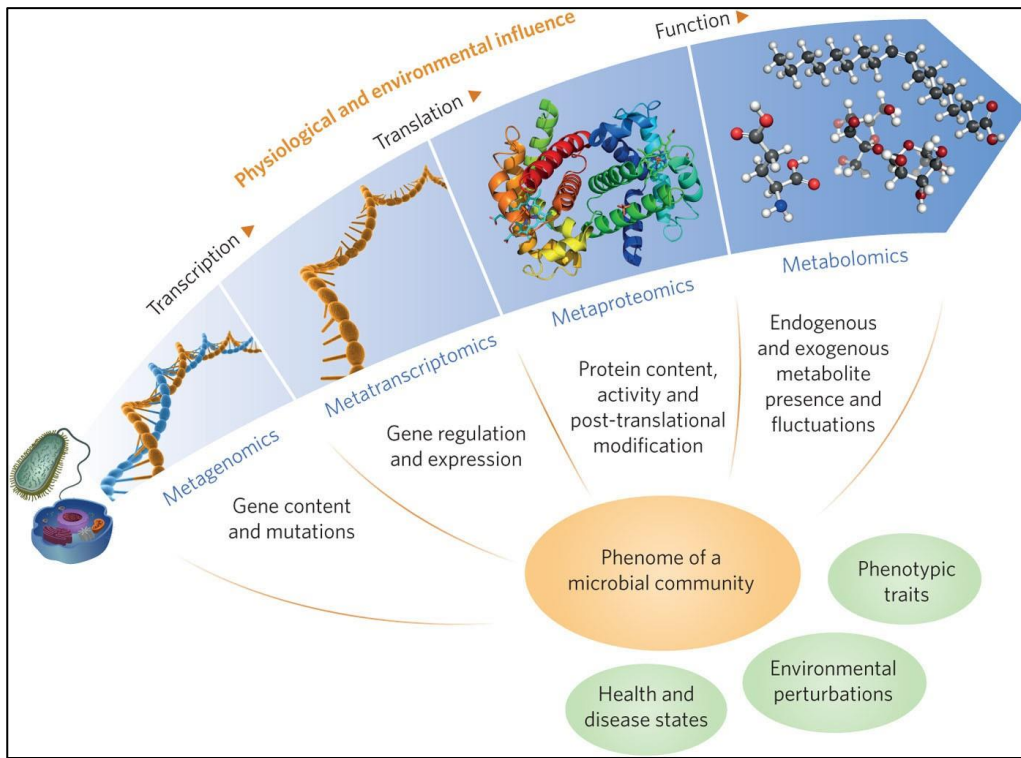


Figure 23: Multi-omics studies for microbiome (Jansson & Baker, 2016)

1.6. Understanding the gut microbiome

The gut contains microorganisms in the range of 10^8 - 10^{11} cells/gram of biomass depending on the region of isolation in the gastrointestinal tract in humans (Sender et al., 2016 (A)). These microbes encode over 3 million genes that generate thousands of metabolites. The species identified in the gut belong to the phyla Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia which are mentioned in figure 24 with some common examples. Among the phyla, Firmicutes and Bacteroidetes compose 90% of the total gut microbiota (Arumugam et al., 2011). The ratio of these two phyla is crucial to differentiating factors of the microbiota and the state of well-being of the host (Magne et al., 2020). The exact type of species and their population can be varied between individuals and within the individuals based on the following factors- geography, diet, age, pharmaceuticals, lifestyle, the birthing process, and infant feeding (Hasan & Yang, 2019). The variable composition of the gut can be considered physiological in the context of healthy gut microbiota concerning age, ethnicity, lifestyle, and dietary habits. At the same time, dysbiosis can have implications leading to disorders. Thus, it is difficult to determine whether the change is beneficial or detrimental.

The association of gut microbiome in dysbiosis can be grouped into intestine disorders and extra-intestinal disorders. Irritable bowel syndrome (Bhattari et al., 2017), Inflammatory bowel disorder (Frank et al., 2007), Celiac disease (Marasco et al., 2016), and Colorectal cancer (Wang et al., 2012) are some of the intestine disorders associated with gut microbiome dysbiosis. Whereas, metabolic disorders like obesity (Ley et al., 2005) and Type 2 Diabetes (Larsen et al., 2010); Central nervous system-related disorders like Alzheimer's (Vogt et al., 2017), Parkinson's Disease (Hopfner et al., 2017), Hepatic Encephalopathy (Bajaj et al., 2012), autism spectrum disorders (Li et al., 2017 (B)) and stress (Bailey et al., 2011) are the lists of extra-intestinal disorders associated to gut microbiome dysbiosis.

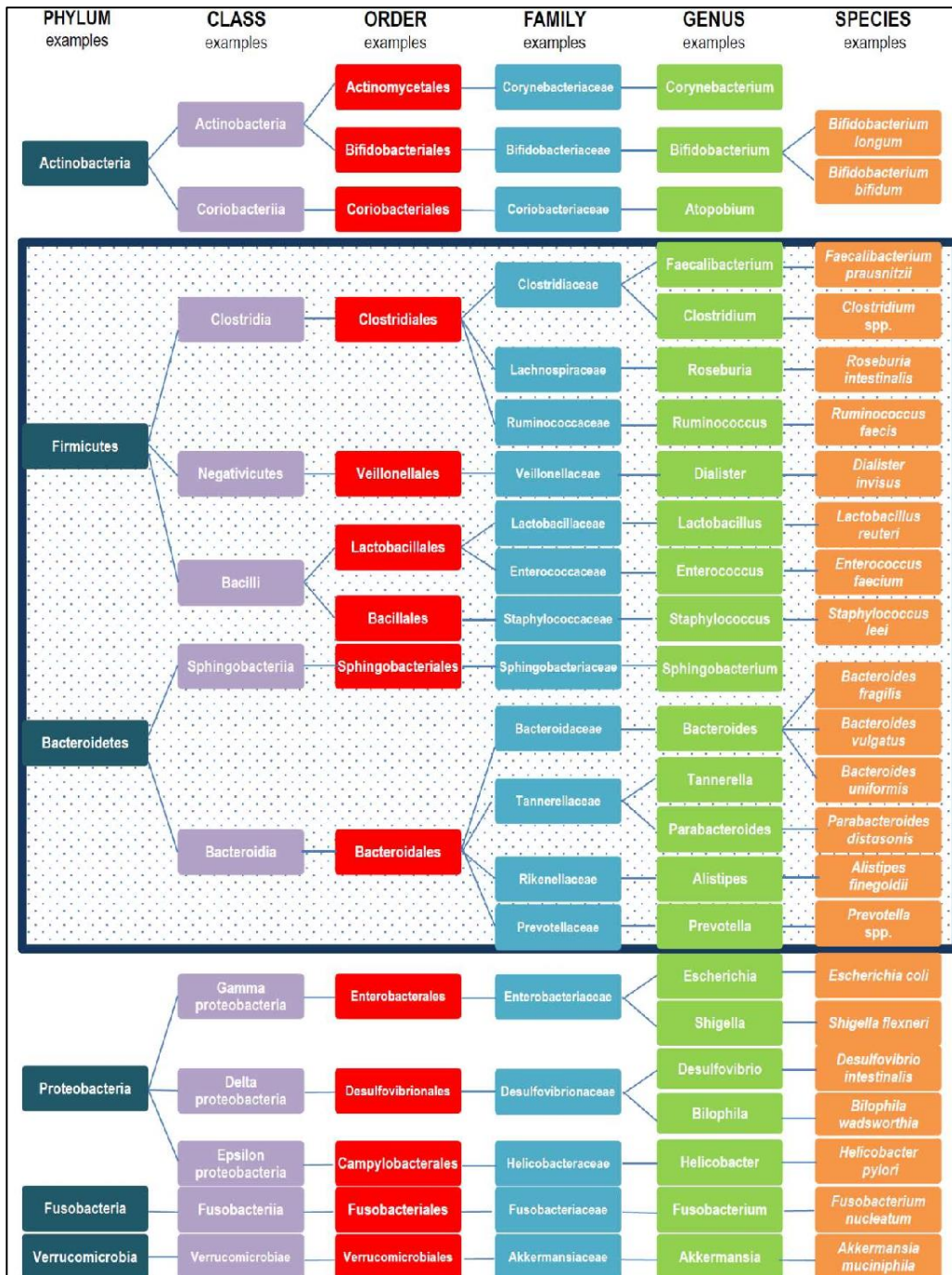


Figure 24: Microbial phylum commonly found in the gut (Rinninella et al., 2019)

The gut microbiome is the single largely studied consortium encompassing metagenomics, metatranscriptomics, metaproteomics, and metabolomics. The reference databases for 4,644 species from the human gut have been established (Almeida et al., 2021), thereby making it an ideal model system for testing new strategies or technologies for microbiomes and omics. The microbiome is usually isolated from the faeces of the host as the specimens are naturally collected, non-invasive, and can be sampled

repeatedly (Tang et al., 2020). Alternatively, several non-human model systems including *in-vitro* models and animal models have been developed to study gut microbiome functions. *In-vitro* models for gut microbiome cultures include short-term batch incubations (De Weirdt et al., 2010), single-stage reactors or semi-continuous systems (Miller & Wolin, 1981), 3D *in-vitro* model (Biagini et al., 2020), multi-compartmental continuous systems like MacFarlane/Gibson three-stage continuous culture system, simulator of the human intestinal microbial ecosystem system, EnteroMix, Lacroix model and TIM-2 (Marzorati et al., 2013). Some of the gut microbes cannot be cultured *in-vitro* owing to their complex growth requirement and thus the usage of animal models can circumvent this problem. Murine models are one of the well-established animal models for gut microbiome studies (Nguyen et al., 2015). They provide several advantages including *in-vivo* manipulations (diet, housing, etc), knowledge of genetics with the availability of genetically modified murine models (knock-outs and disease models), a low-cost, omnivorous mammal with gut physiology comparable to humans, and provide homogenous genetic background (in-bred). Thus, in this present study, I will evaluate the performance of the microfluidic technology with an *in-vitro* species mix reproduced from the gut microbiome and murine model for the gut microbiome.

1.7. Scope of the thesis

In the previous work, the ChipFilter design and usage in proteomic sample preparation were established. The current works described in the thesis will try to advance the application of microfluidic ChipFilter to process the microbial cells to allow sample preparation specific for metaproteomic applications. As described in figure 25, the development will be focussed on processing several microbial models/sample systems (3 model species mixture, gut standard, and murine gut microbiome) with the ChipFilter for proteomics pre-analytical sample preparation and beyond like nucleic acid isolation or microbial sample collection.

Chapter 1 introduces the different microbiomes identified, methods of studying microbiomes (metaproteomics and metagenomics) as well as the challenges, and recent developments in microfluidic technology with a focus on microbial cell processing for proteomic or genomic applications. Further, advancements in multi-omics approaches and gut microbiome models that will be useful to establish the advancements made with the ChipFilter are discussed.

Chapter 2 demonstrates the construction of ChipFilter along with five different workflows that will be useful to process microbial samples for different applications and that will be used in subsequent chapters. Next, new modifications made to the workflow from one used in the previous study are explained with an elucidation of single microbial cell type processing. Finally, a comparison between the two workflows is outlined to show their differences and commonalities.

Chapter 3 will compare the microbial proteomic sample processing between ChipFilter and other existing methods. Sample complexity is made with mixed cell communities of three model species varying in phenotype and cell numbers. The objective here will be to demonstrate that ChipFilter workflows are on-par or better than the current methods available for proteomic sample preparation.

Chapter 4 provides the proof of concept by showing ChipFilter sample processing of microbiome isolated from murine gut for proteomic application. Bioinformatic pipelines used for protein sequence database generation are shown. Finally, the results showing protein identification and taxonomy distribution are given.

Chapter 5 demonstrates the capacity of the ChipFilter workflow to isolate nucleic acids after the processing of cells for proteomic application. The nucleic acids isolation workflow was studied for their characteristics and validated by polymerase chain amplification.

Chapter 6 combines the proteomic and nucleic acid sample preparation process with ChipFilter to allow for multi-omics sample processing is shown. This is demonstrated with a complex mixture of twenty-one gut microbial strains with different cell abundance and taxonomy.

Chapter 7 proposes the usage of ChipFilter for the collection of bioaerosols with or without modification to the original design. The stability of the ChipFilter structures is demonstrated and application examples are shown.

Chapter 8 finally draws the general conclusions and perspectives of my thesis work.

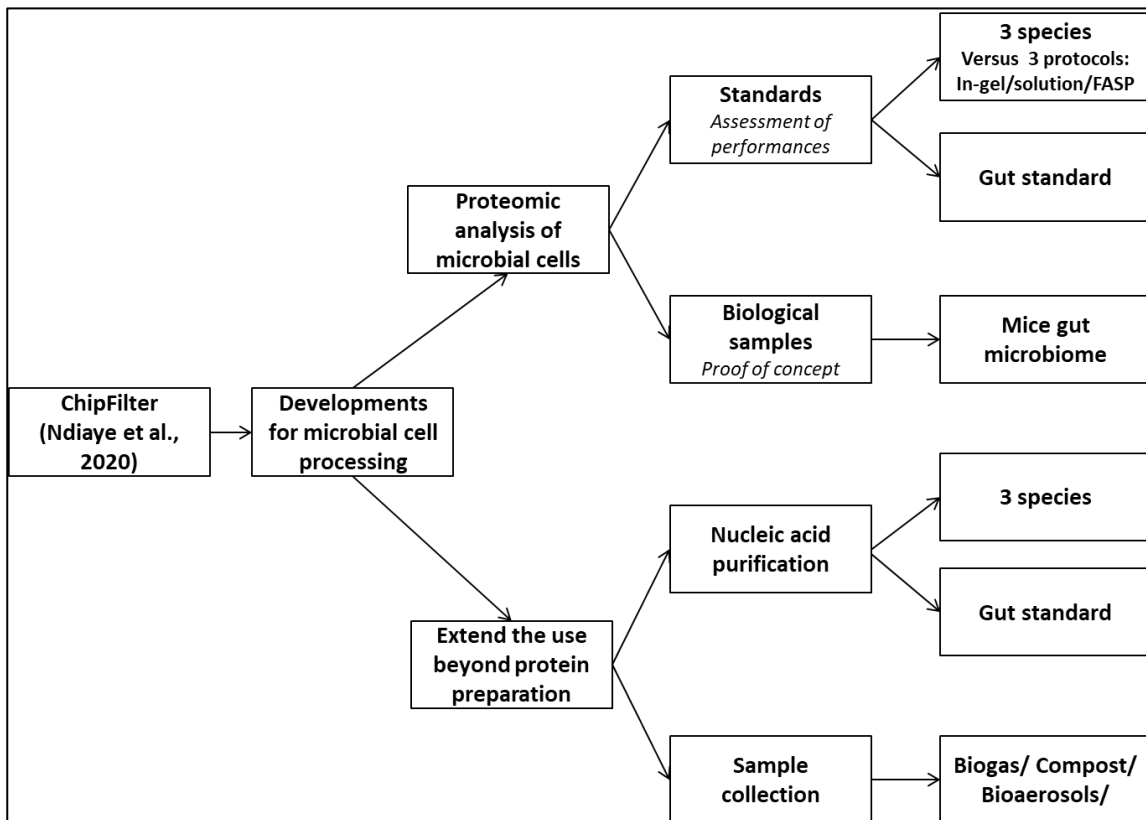


Figure 25: Overall organisation of the thesis work

2. Development of ChipFilter **workflows for whole-cell microbial** **sample processing**

2.1. Introduction

The ChipFilter device developed in the lab can perform cell lysis, protein concentration, and rapid enzymatic treatment of proteins from eukaryotic cells (yeast or human). It was shown to obtain better proteome coverage than the FASP protocol (NDiaye et al., 2020). These results allowed us to envisage the use of ChipFilter for the analysis of microbial cells.

However, before putting it to use on microbes, different scenarios have to be addressed. At first, is to achieve cell lysis inside ChipFilter. The conditions used must be effective for cell walls that are superior to the plasma membrane of eukaryotic cells. Next, to effectively recover most of the peptides and to avoid inter-sample contamination between different microbial samples.

In this chapter, I will address these questions. The ChipFilter design that was standardized in the laboratory (NDiaye et al., 2020) was fabricated using necessary microfluidic technologies. Next, the use of ChipFilter for microbial sample preparation will be optimized by establishing specific workflows that can be used for diverse microbial species (bacteria and fungi). Additionally, the conditions to perform on-chip microbial cell lysis and protein extraction are discussed with all the considered workflows. The discussion of workflow standardization extends to overcoming the problem encountered with proteins/cell debris retention in channels of the Fluigent pump systems and capillaries used during cell loading. Complete sample preparation for bottom-up proteomics was tested for independent cultures of *Escherichia coli*, *Saccharomyces cerevisiae*, and *Bacillus subtilis* to identify their respective peptides and proteins. Differences arising in the identification of proteins while using different peptide elution workflows with ChipFilter are explained. In general, this chapter will explain the different ChipFilter workflows to be used in the study for microbial sample preparation and adaptations made to meet the demands of metaproteomic sample preparation.

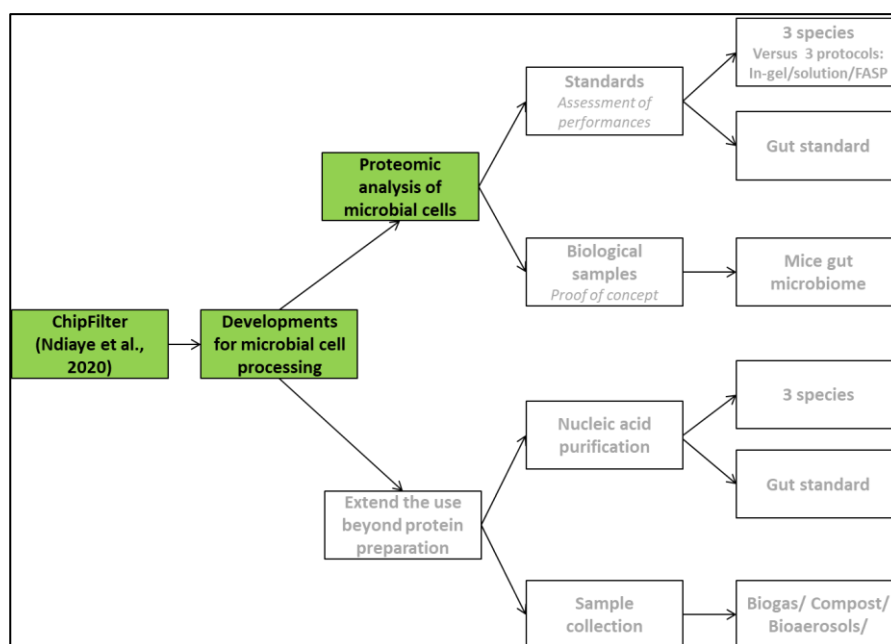


Figure 26: Steps discussed in chapter 2

2.2. Material and methods

2.1.1. Biological material

Single microbial colonies of *Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacillus subtilis* were used. Each of these model cells represents the fungi, Gram-negative, and Gram-positive bacteria types respectively. Luria-Bertani (LB) agar (Thermo Fisher Scientific) plates were inoculated with the cells and incubated at 37 °C overnight. Inoculum into LB broth was made and cells were cultivated until the optical density reached 1 at 600 nm. Before harvesting, cells were counted using a glass slide (Fisher Scientific) under a microscope. Collection of the cells was done by centrifugation at 2400 g for 5 minutes at room temperature. A single wash with phosphate-buffered saline (PBS) pH 7.4 was performed before pelleting and storage at -80 °C until further use. Appropriate dilutions were made with 50 mM ABC buffer to obtain the necessary cell density.

2.1.2. ChipFilter development

The ChipFilter device was fabricated at the Institut Pierre-Gilles de Gennes (IPGG) platform (Paris). The ChipFilter devices were developed in three stages involving (i) the development of mould by photolithography, (ii) die-casting replication of the mould designs onto PDMS, and (iii) molecular filtration membrane insertion and assembly. The summary of these steps can be understood in figure 16. For the present work, the same design conditions were used and they are explained below.

Photolithography

Photolithography is the most widely used method that consists of transferring patterns onto a layer of photosensitive resin (called photoresist). The photoresist is spread evenly on a smooth support called a wafer and then exposed to light through an optical mask. The optical mask is a transparent support on which the patterns are printed in opaque (Iron or Chrome oxide). The purpose of using the mask is to protect part of the photoresist layer from incident radiation. A beam of electrons or photons from different sources of radiation can be used in photolithography. Once exposed to the energy source, the polymerization of the photoresist is activated or inhibited (negative or positive resin). It uses radiation wavelengths ranging between 300 and 450 nm to have precision in the sub-micrometre. The removal of unpolymerized parts is the last step before transferring patterns onto the solid support like silica. The use of short wavelengths improved the accuracy of the technique.

- **Drawing of the patterns:** Computer Assisted Design software package called CleWin (CleWin5) was used to draw the patterns that were printed on a transparent sheet using a photoplotter Filmstar-Plus equipped with a 670 nm Laser-diode (BungardElektronik) with resolutions of 5 μm (object) and 18 μm (space between objects). Printing was done with a vertical resolution of 25, 400 dpi (dot per inch) and a horizontal resolution of 8128. Two sets of masks indicated by 1 and 2 in figure 27 are necessary for a single silicon wafer for one entrance of the ChipFilter. Two pairs of wafers were made capable of fabricating four ChipFilter in a single cast.
- **Manufacture of the mould:** Silicon wafer was coated first with a resin layer consisting of negative resin SU-2007 spread by rotation (spin coating) at a speed of 2000 rpm for 15 seconds to obtain a thickness of 8 μm . After spreading the first layer on the wafer, the assembly is heated to 95 °C for 2 minutes. The resin is then exposed to UV light with mask 1 using a UV-KUB2 aligner (LED, 23.4 mW/cm², 120 mJ/cm², 5 seconds). The patterns of mask 1 are thus transferred to the mould which will again be at 95 °C for 2 minutes. Then, the second layer of resin SU-2050 is spread by rotation at 4080 rpm for a thickness of 50 μm above the first. Mask 2 is aligned with the patterns already visible on the wafer using UV-KUB3 aligner. The wafer, now covered with layers of resin, is exposed to UV light (LED, 40 mW/cm², 160 mJ/cm², 4 seconds). The pattern is heated a second time to fix all transferred designs. Finally, the wafer is immersed in a solution provided by the resin called the developer whose role is to dissolve all the parts of the resin that have not been exposed to UV light. We then obtain a 3D mould resulting from the patterns of the 2 masks.

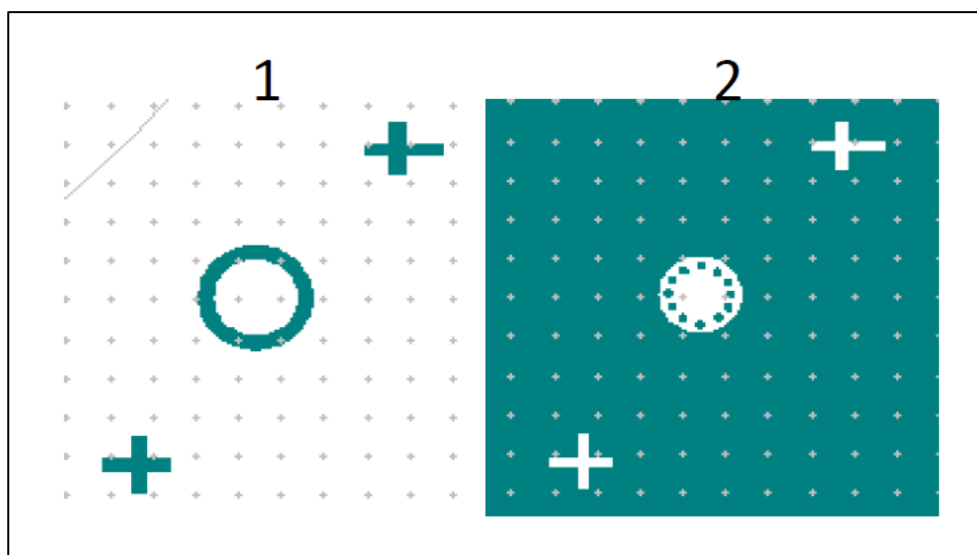


Figure 27: Diagram of two masks (1 and 2) necessary to create the mould for ChipFilter

Die-cast replication

Replication moulding is done by deposition of a polymer heat-crosslinking agent (PDMS) on the mould. The patterns of the 3D mould are reproduced by cast replication. For this, base and cross-linking agent (Sylgard 181, Dow Chemical) in the ratio of 10:1 was mixed and added to the mould which corresponds to approximately 40 g in weight. The mixture is then degassed and deposited on the mould. The assembly is placed in an oven at 70 °C for at least one hour for PDMS crosslinking. After cross-linking, the solid structure formed by the polymer will reproduce the negative of the mould.

Molecular filtration membrane insertion and assembly

After cross-linking, demoulding, and cutting the PDMS containing reaction chambers and alignment marks as shown in figure 28. The entrance to each reaction chamber was made by punching a hole with a sampling tool (Rapid core, WellTech). To achieve a tight connection between the ChipFilter and the microfluidic pump and the LC, the diameter used when punching was made slightly smaller than the outer diameter of the capillaries. The elasticity of PDMS cross-linked allows this watertight connection to be obtained without the use of adhesive products. The punching was performed before processing the PDMS with plasma (20 W, 8 sccm O₂ flow, and 0.13 mbar pressure for 1 min). To enable the adhesion of the two sides of the reaction chamber oxygen plasma treatment was done to make the surface hydrophilic and enable crosslinking by the formation of covalent bonds. Before assembly, the filtration membrane is placed on the area provided. Covering of the reaction chamber was ensured by aligning the marks patterned to ensure the superposition of the two reaction chambers visualising under a microscope.

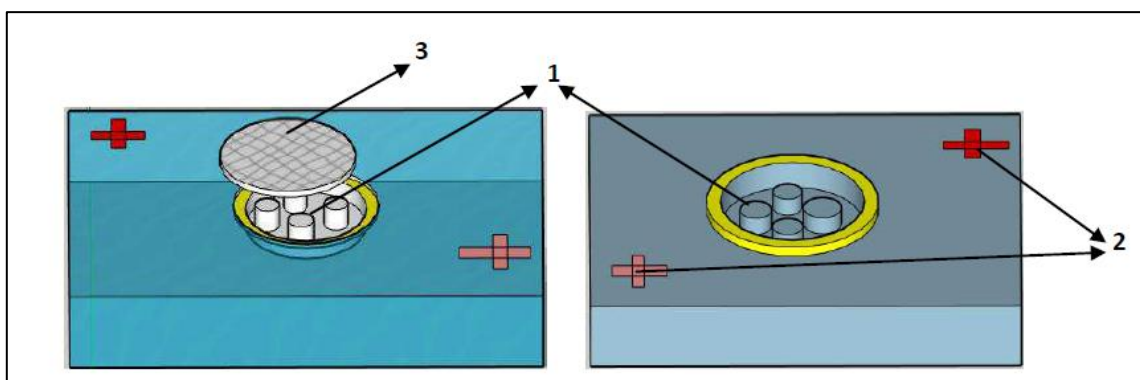


Figure 28: The structures after PDMS mould replication. 1, 2, and 3 indicate pillar structures, alignment marks, and filtration membrane

2.2.3. Workflows for sample processing

Upon fabrication of the ChipFilter, the microfluidic system is to be used for accommodating different chemical reactions that can allow cell lysis, clean-up, protein unfolding, proteolysis, and peptide recovery. Furthermore, the nucleic acid components retained in the reaction chamber will also be removed for studies. The major steps can be found in figure 29 and explained below in detail. They differ mainly in the connections (WF1-4) and the use of lysozyme in the buffers (WF5). A summary table mentioning the important difference between workflows can be found in the annex.

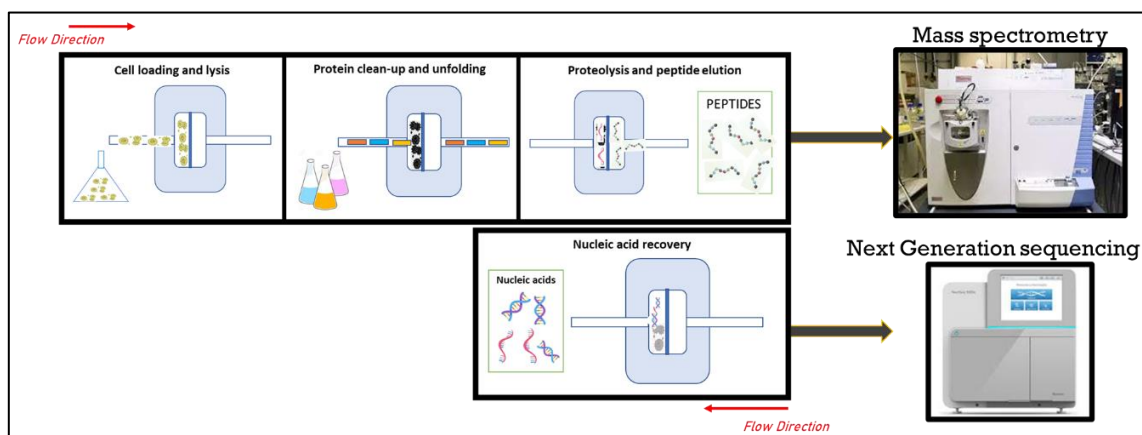


Figure 29: The key steps involved in sample preparation using ChipFilter

To enable automation in sample processing and movement of liquids through the microchannels, external apparatus like a pump with flow controller (Fluigent), P-cap reservoir (Fluigent), M-switch valve (Fluigent), and computer-based automation software (Fluigent) to control the feedback from the apparatus was utilized. Additionally, a piston syringe (Agilent) and syringe pump (Harvard Apparatus) were used to introduce cells inside the ChipFilter. The process for introducing the cells into the device was done using

both the Fluigent system and the syringe piston, an explanation for this can be found in section 2.3.

Pump and flow controller

An external pump is necessary to tackle excessive hydrodynamic resistance of the membrane of molecular filtration to make it compatible with capillary actions and electro-osmosis modes. Also, it is needed to control pressure and flow rates during the sample preparation to ensure the integrity of the ChipFilter. Monitoring the flow rate also makes it possible to calculate the amount of sample or reagent added to determine the reaction time. During this study, we used a Fluigent MFCS pump and Flow-EZ flow controller. The Flow-EZ pump is associated with a Flow- Unit-S that was also used. The pressure range of the pump is 0 to 1000 mbar while the flow controller operates between 0 and 8 $\mu\text{l}/\text{minute}$.

P-Cap Reservoir

P-Caps are accessories for holding the standard tubes (1.5; 2; 10; or 50 mL) useful as samples or reagent reservoirs connected to a microfluidic device. They allow the pressurization of these tubes.

M-switch valve

The M-switch is a multidirectional valve, with ten inputs and one output. This configuration makes it possible to operate injections sequentially from ten different tanks. In this case, injections to be made from samples, chemical reagents, enzymes, and rinse solution tubes can be automatically done.

Workflow 1 (WF1)– Capture of digested peptides directly in the injection loop of LC systems:

- **To load the cells:** Fixed volume of the cells suspended in an appropriate buffer like 50 mM ABC was loaded into the device at a flow rate of 2 $\mu\text{l}/\text{min}$ for the Fluigent system and 0.01 ml/min for the piston pump.
- **To wash the cells:** (optional) Pass 20 μl at the rate of 2 $\mu\text{l}/\text{min}$ of 50 mM ABC to remove debris and make the cells in contact with the device.
- **Cell lysis:** This is done to lyse the cells. Use appropriate lysis buffer (1 % (w/v) Octyl β -D-glucopyranoside (ODG), 150 mM Tris-HCl pH 8.8, protease inhibitor) for 45 μl at the rate of 2 $\mu\text{l}/\text{min}$ and 30 μl at the rate of 1 $\mu\text{l}/\text{min}$.
- **DTT Reduction:** This is performed to reduce the cysteine residues and it is done by using 20 mM DTT in lysis buffer without protease inhibitors or lysozyme for 45 μl at the rate of 2 $\mu\text{l}/\text{min}$ and 30 μl at the rate of 1 $\mu\text{l}/\text{min}$.

- **IAA Alkylation**: This is performed to alkylate the reduced cysteine residues and it is done by using 50 mM IAA in lysis buffer without protease inhibitors or lysozyme for 45 μl at the rate of 2 $\mu\text{l}/\text{min}$ and 30 μl at the rate of 1 $\mu\text{l}/\text{min}$.
- **Washing**: This is performed to remove the excess reagents and it is done by using 50 mM ABC buffer for 45 μl at the rate of 2 $\mu\text{l}/\text{min}$ and 30 μl at the rate of 1 $\mu\text{l}/\text{min}$.
- At this point, if the device exit is connected to the LC system for direct capture of peptides, make sure the Loading pump flow is set to 5 $\mu\text{l}/\text{min}$ with 99% buffer A and 1% buffer B. Also, keep the LC in “LOAD” mode (the trapping column is disconnected from the sample loop). The LC system should be in 99% A at least 15 minutes before the proteolysis and peptide carrying in the injection loop.
- **Proteolysis**: This is performed to generate peptides and it is done by using 2 μg of trypsin (Promega) in 50 mM ABC buffer for 20 μl at the rate of 2 $\mu\text{l}/\text{min}$.
- **Digestion**: To facilitate digestion, pass 50mM ABC buffer of 20 μl at the rate of 2 $\mu\text{l}/\text{min}$ and 50 μl at the rate of 0.5 $\mu\text{l}/\text{min}$.
- **Elution**: To facilitate elution, pass 50 mM ABC buffer of 15 μl at the rate of 1 $\mu\text{l}/\text{min}$ and 30 μl at the rate of 2 $\mu\text{l}/\text{min}$.
- Finally, connect the precolumn by changing the mode to inject, make the flow rate 10 $\mu\text{l}/\text{min}$, and perform this step for 20 minutes with 99% buffer A and 1% buffer B to transfer the peptides to the trapping column.
- Perform LC-MS/MS using suitable workflows for the instruments.

Workflow 2 (WF2) – Capture of digested peptides in a vial:

- **To load the cells**: Same as in WF1.
- **To wash the cells**: (optional) Same as in WF 1.
- **Cell lysis**: Same as in WF 1.
- **DTT Reduction**: Same as in WF 1.
- **IAA Alkylation**: Same as in WF1.
- **Washing**: Same as in WF 1.
- **Proteolysis**: Same as in WF 1.
- A 1.5 ml centrifuge tube is placed at the exit of the ChipFilter to collect the flowing solution.
- **Digestion**: Same as in WF 1.
- **Elution**: Same as in WF1.
- Remove the tube collecting the liquid and dry it in a vacuum drier. Reconstitute the dried peptides in an appropriate volume of 0.1% TFA and inject it into an LC.
- Perform LC-MS/MS using suitable workflows for the instruments.

Workflow 3 (WF3) – Shorter digestion time for one hour without applying fluid flow:

- To load the cells: Same as in WF1.
- To wash the cells: (optional) Same as in WF1.
- Cell lysis: Same as in WF1.
- DTT Reduction: Same as in WF1.
- IAA Alkylation: Same as in WF1.
- Washing: Same as in WF1.
- Proteolysis: This is performed to generate peptides and it is done by using 2 µg of trypsin (Promega) in 50 mM ABC buffer for 10 µl at the rate of 2 µl/min.
- Digestion: To facilitate digestion, pass 50 mM ABC buffer of 40 µl at the rate of 2 µl/min.
- A 1.5 ml centrifuge tube is placed at the exit of the ChipFilter to collect the flowing solution.
- Elution: To facilitate elution, pass 50 mM ABC buffer of 15 µl at the rate of 1 µl/min and 30 µl at the rate of 2 µl/min.
- Remove the tube collecting the liquid and dry it in a vacuum drier. Reconstitute the dried peptides in an appropriate volume of 0.1% TFA and inject it into an LC.
- Perform LC-MS/MS using suitable workflows for the instruments.

Workflow 4 (WF4) – Nucleic acid recovery after peptide elution:

- All the steps until peptide elution were performed either by WF1, 2, or 3.
- The nucleic acid recovery solution was either Milli-Q water or 50 mM ABC buffer for the appropriate volume.
- The fluid flow inside the device must be changed to the opposite side from the direction used until peptide elution.
- A 1.5 ml centrifuge tube is placed at the exit of the ChipFilter to collect the flowing solution. A flow rate of 2 µl/min for an elution volume ranging from 30 µl – 45 µl. Double elution was performed as optional.
- The collected solution is quantified and used for appropriate analysis.

Workflow 5 (WF5) – Addition of lysozyme to enhance cell lysis of Gram +ve bacteria:

- Any workflow described above can be used with the addition of lysozyme at a concentration of 50 ng/ml lysis buffer. This was only used for the gut standard microbiota described in chapter 6. 2.2.3

2.2.4. LC-MS/MS analysis and protein identification

Samples were analysed by nanoLC-MS/MS in DDA high-energy c-trap dissociation mode using an RSLCnano UltiMate™ 3000 System coupled to a nanoESI Q-Exactive or Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). For the ChipFilter method, the peptides recovered in the trap column were separated on a capillary reverse-phase C18 column Pepmap 75 μm i.d. \times 50 cm length (ThermoFisher Scientific) at 45°C with a linear 120 min gradient elution from 2.5% to 60% of buffer B (water/acetonitrile/formic acid 10%: 90%: 0.1% (v/v/v)) in buffer A (water/acetonitrile/formic acid 98%: 2%: 0.1% (v/v/v)) at a fixed flow rate of 220 nl/min. The dried peptides were resuspended in 7 μl of 0.1 % TFA solution (v/v) and 6 μl was used for single-shot injection. Trapping was done with C18 Pepmap 300 μm i.d. \times 5 mm length column (Thermo Fisher Scientific) and analysed in nanoLC MS/MS with a 120 minutes gradient as described earlier.

Sample analysis was performed with a Q-Exactive mass spectrometer operated in nanoESI at 1.7 kV. Full MS survey scans were recorded over the m/z range of 400 – 2000 with a resolution of 70,000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 100 ms. Up to 15 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 30, with a precursor isolation window at 2 m/z, resolution set at 17,500 with AGC value at 1E5 with a maximum injection time of 120 milliseconds. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 seconds.

The spectra were analysed by Proteome Discoverer 2.4 (Thermo Fisher Scientific) software package. The mascot search engine (Matrix Science Mascot 2.2.04) was used for database search with the SwissProt database with taxonomy specific for the cell type mentioned. The peptide precursor mass tolerance was 10 ppm and the fragment mass tolerance was 0.02 Da. Trypsin was selected as the hydrolytic enzyme. The allowed maximum missed cleavage was two, with modifications Carbamidomethylation (C), Deamidation (NQ), and Oxidation (M). Proteins with at least one high-confidence peptide and six amino acids were validated. Target FDR was set at 0.01. Proteins were filtered for “Master” only and “High” Protein FDR Confidence Combined.

2.3. Loading and lysis of microbial cells into the device

One of the most important differences in the workflow between the earlier study and the present work with the ChipFilter is in how the microbial cells have been introduced into the system. Earlier work mainly focused on the usage of cellular lysates that was introduced to the ChipFilter by using the Fluigent system. One of the P-Cap was used to introduce the lysate or in some cases the whole cells, followed by other reagents

from the same position or adjacent caps. In the previous study, experiments using whole cells were limited and mainly focussed on a single species (Human cancer cell line).

A key factor in this study is the usage of lysis buffer-only mediated (no pre-treatment) cell lysis inside the ChipFilter. Studies have shown that (Fradique et al., 2020) using commercial lysis buffer (B-Per) containing non-ionic detergent proved to be highly useful for cell lysis of *Escherichia coli*. Also, indicate the latent activity of lysozyme on Gram-positive bacteria to induce cell lysis (Mahalanabis et al., 2009). In the previous work with ChipFilter, the authors used the lysate of *Saccharomyces cerevisiae* after bead-beating (mechanical) pre-treatment to perform proteolysis. In this study, for the first time, I test the lysis of whole cells of bacteria (Gram-positive and negative) and fungi within the ChipFilter. Additionally, the inclusion of lysozyme was made to the lysis buffer during the cell lysis step for the gut microbiome standard that consisted of 21 strains composing 18 bacterial strains, 2 fungal strains, and 1 archaeal strain.

When the previous workflow developed for whole human cell lysis was used for microbial cells in ChipFilter, I observed a protein carry-over effect which resulted in the identification of proteins from the previous experiment to be found in the running experiment. This was also confirmed with the blank experiment using WF1 where there was no cellular loading, yet identification of a significant number of proteins. Hence, these proteins were termed “contaminants”. Furthermore, a protein identified from relics or human/eukaryotic origin in pure microbial cultures was also identified. It is imperative to note the percentage of contaminants was greater when the target cell number was lowered (figure 30). It can be identified only when the protein search engine included all taxonomy databases during MS spectral data treatment.

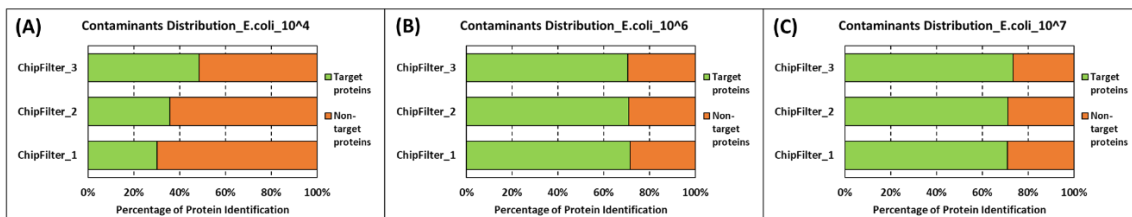


Figure 30: The decrease in contaminants with increasing target cell number 10^4 (A), 10^6 (B), and 10^7 (C). Percentage of the target (Green) and non-target (orange) proteins identified

This situation led to several hypotheses regarding the origin of the contamination effect that includes suspicion about (i) cells retained inside the capillary or other connections that enable fluid flow, (ii) sample loop or trapping column, and (iii) originating during device fabrication. To overcome the possibilities, different washing strategies were undertaken. This included rinsing the Fluigent system and device (back and forward flow)

with 50% Acetonitrile (ACN), 100% Methanol, and Milli-Q water before cell loading/sample preparation. The LC systems with trapping column were also washed with 100% B buffer (water/acetonitrile/formic acid 10%: 90%: 0.1% (v/v/v)) for at least 30 minutes before trapping. It was only possible to reduce the contaminants by under 10% with these methods, but the origin of the problem was identified in the cell loading procedure using Fluigent systems.

I believe that even though the tubing used is made of polyether ether ketone (PEEK) that can be resistant to fouling, the possibility of it not occurring is scarce. Hence, I conclude that the fouling was the cause of carryover proteins. To overcome this problem, I devised a strategy to replace all the capillaries with new capillaries of the same material at a regular interval (every 6 months) and use an alternative strategy to introduce cells into the ChipFilter. This led to the usage of a piston syringe and pump to introduce cells or lysate. The blanks after using this strategy gave only a few tens of proteins as compared few hundred proteins identified in blanks before using this procedure (data not provided). This strategy could also allow the transportation of samples collected on-field in the ChipFilter, where the sample collection and processing are made at different times and places as discussed in chapter 7.

2.4. Single microbial sample preparation and identification

The ability of the ChipFilter to allow the preparation of microbial samples for proteomic analysis was evaluated on single microbial cultures. Approximately one million *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* were loaded onto the device. The WF1 was used for sample preparation, and the peptides were analysed and identified as previously described.

The number of proteins identified for *E. coli*, *B. subtilis*, and *S. cerevisiae* is $1\,121 \pm 58$, 824 ± 50 and $1\,339 \pm 44$ respectively; whereas the number of peptides was $4\,694 \pm 340$, $5\,802 \pm 463$ and $7\,959 \pm 412$ respectively. The number of peptide spectral matches (PSMs) identified include $39\,466 \pm 1\,757$, $36\,302 \pm 1\,083$, and $75\,096 \pm 1\,120$ respectively as shown in figure 31. The identifications of the peptides specific to the species prove that bacterial and fungal cells can be lysed without pre-treatment by the ChipFilter process.

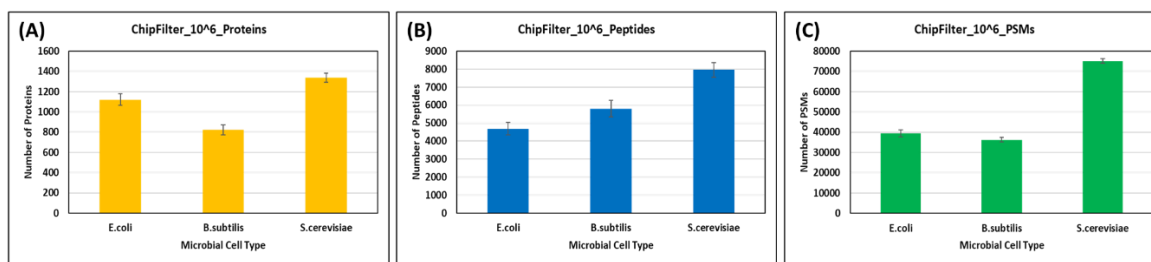


Figure 31: The proteins (orange), peptides (blue), and PSMs (green) identified for *E. coli*, *B. subtilis*, and *S. cerevisiae* using ChipFilter

As per literature evidence, *E. coli* cell lysis with RapiGest and SDS (2% m/v) based buffers with heating pre-treatment at 95 °C for 30 minutes allowed the identification of fewer than 1 000 proteins, 4 000 peptides, and 10 000 PSMs (Tanca et al., 2013). In comparison, the ChipFilter method identified a similar number of proteins and peptides or a higher number of PSMs. It must be noted here that the number of cells used in ChipFilter workflow is 1000-fold lesser than the mentioned methods.

B. subtilis is gram-positive bacteria with a peptidoglycan cell wall that is usually difficult to lyse as compared to other bacterial cell types. This makes it necessary to use detergents capable of penetrating this cell wall and also may require enzymatic treatments (discussed in a later section). Unlike SDS-based detergents that require pre-treatment for gram-positive bacterial lysis, ChipFilter workflow with ODG-based buffer proved effective in cell lysis and subsequent peptide identification. Literature evidence for intact-spore extracts and vegetative cells of *B. subtilis* identified 1 428 and 1 258 proteins respectively which is more than the proteins identified by the ChipFilter method (Swarge et al., 2018). In this study, the usage of urea buffer with bead-beating pre-treatment was demonstrated and it can be interesting to incorporate the urea buffer in future ChipFilter studies.

Fungal cells like *S. cerevisiae* are eukaryotic having a defined nuclear membrane and cell wall. They are tough and are known to be difficult to lyse completely due to these structures. Fortunately, the ChipFilter method can lyse whole cells of *S. cerevisiae* without the need for pre-treatment. Earlier studies of the yeast lysate with the ChipFilter method identified 1 815 proteins and 10 201 peptides. When using whole cells of the same species the number of identifications was lowered. It is evident that the processing of lysate was efficient, but the capacity to automate cell lysis can be advantageous too. Also, the MS instruments used in both studies are different with the one used in the lysate study being superior.

2.5. Peptide elution: Trapping column (WF1) versus vial (WF2)

The advantage of WF1 is in complete automation of the process of sample preparation which in the future may offer the possibility of a direct ionization for LC-MS/MS studies. However, the elution volume is limited by the capacity of the sample collection capillary in the LC and, as seen in some of my experiments, the capillary is prone to contamination by peptides already eluted from the previous experiment. I have overcome this problem by washing with organic solvents like 50% acetonitrile or 100% isopropanol, but it requires regular maintenance.

To overcome these limitations, I proposed a new workflow (WF2). In the case of WF2, the final step before injection into the LC-MS/MS requires human effort to collect and process the peptides before injection. This step allows us to increase the elution volume and collect the peptides from the beginning of the digestion process. Additionally, fractionation of the peptides to enable multi-dimensional LC-MS/MS can be performed externally. Opportunity is there to quantify and clean the peptides before injection. Although these possibilities have not been tested, the scope is still available to expand the usage of the ChipFilter.

To compare the WF1 and WF2 in terms of their protein identification and the nature of peptides eluted, a mix containing one million cells from the three microbial species with the cell number of 3×10^4 (mix 2, preparation described in section 3.1) was used, and the resulting peptides were analysed by LC-MS/MS as described. The experiments were performed in triplicates.

In terms of protein identification, the WF1 identified more proteins ($1\,226 \pm 189$) than WF2 (982 ± 166) with no statistical significance (t-test) as shown in figure 30 (A). The difference is not very high but it is to be noted that the loss of proteins is still higher by workflow 2 as it requires additional processing steps before LC-MS analysis. The overall distribution of proteins between the methods however, showed 65.1% (1 322) proteins to be identified in both methods, but 25.6% (519) proteins were identified only by WF1 and 9.4% (190) only by WF2 is shown in figure 30 (B). While the number of proteins slightly differed, the translation of this into their distribution for specific species remained indifferent for *E. coli*, *S. cerevisiae*, and *B. subtilis* is shown in figure 32 (C and D). This indicates that the peptide elution procedure did not affect the species-level protein identification.

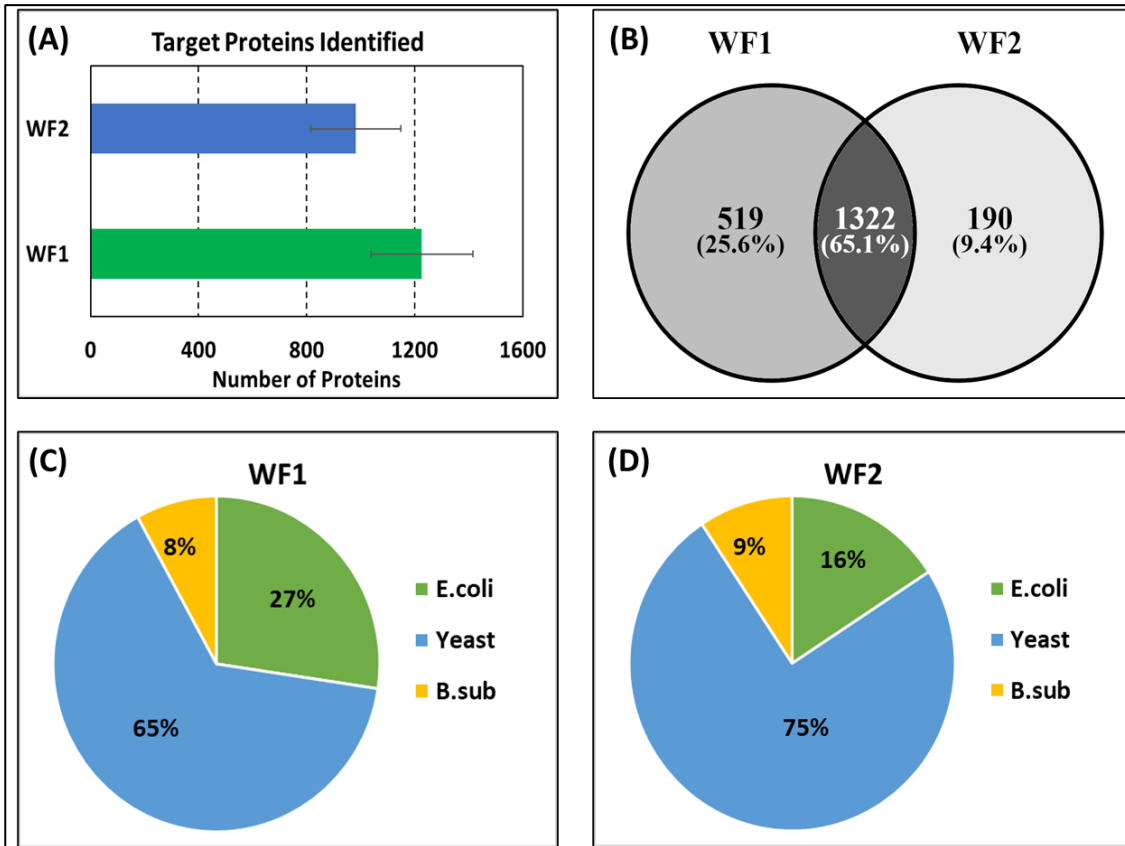


Figure 32: The identified target proteins (A), distribution of target proteins between each other (B), and species (C, D) for ChipFilter WF1 and WF2

Furthermore, the differences like peptides eluted by the two workflows were explored. First, the number of peptides and PSMs identified by each workflow was compared (figure 33). The number of peptides identified by WF1 ($6\,385 \pm 966$) was higher than WF2 ($4\,462 \pm 904$) but without statistical significance as seen for proteins. A similar result was obtained for PSMs as well with WF1 ($43\,202 \pm 3\,095$) generating higher PSMs than WF2 ($39\,593 \pm 3\,984$). An interesting result was observed when the distribution of peptides was studied between the workflows. Peptides identified only by WF2 were 2 088 (17.6%) while that identified by WF1 were twice 4 511 (38%). The peptides common between the two workflows were slightly higher at 5 268 (44.4%). It is important to note that the type of peptides collected by each method is unique even though they correspond to the same protein as established by the recent work (Van Den Bossche et al., 2021). This difference in peptide distribution must be understood and hence nature of the peptides generated was studied.

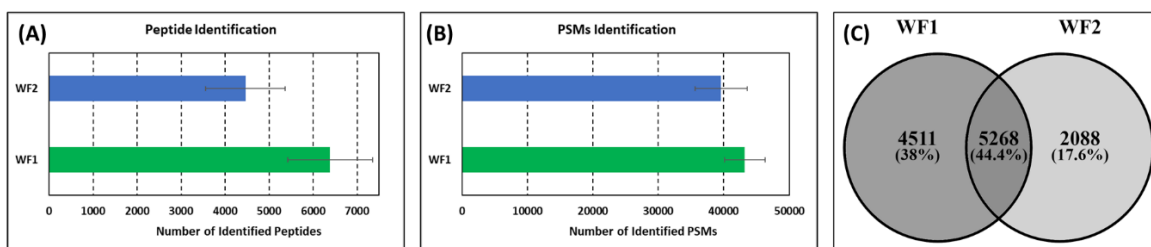


Figure 33: The identified total peptides (A), PSMs (B), and distribution of peptides (C) between for ChipFilter WF1 and WF2

Probing into the physiochemical properties like missed cleavages, peptide length, theoretical mass, and hydrophobicity was essential to further understand the variation generated (figure 34). The missed cleavages generated in the peptides during the different workflows showed that the percentage of missed cleavages was slightly higher for WF1. This can be explained by the fact that it has higher peptide identification. Additionally, the fraction of peptides identified in WF1 consists of only the eluted peptides, while WF2 consists of peptides removed during the digestion process in addition to the eluted peptides. This can as well be seen in terms of the peptide length and positive ion mass distribution, where the peptides shorter than 10 amino acids and 1000 Da respectively were significantly greater (p-value 0.012) in WF2 than in WF1. It can be possible that the collection of small peptides by WF2 can contribute to diversity in the peptide distribution seen earlier. The hydrophobicity index distribution was unaffected between the workflows.

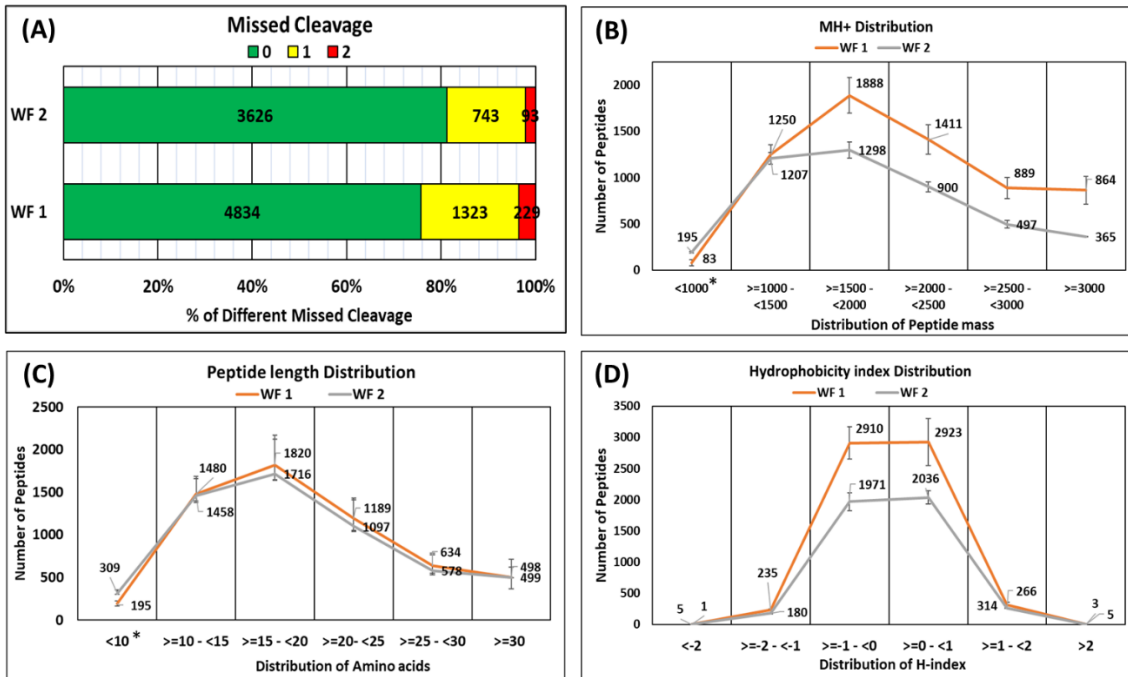


Figure 34: The physicochemical properties like missed cleavages (A), peptide length (B), theoretical positive ion mass (C), and hydrophobicity (D) for ChipFilter WF1 and WF2

In conclusion, the WF1 and WF2 have their advantages and limitations during operation. Additionally, they elute completely different-sized peptides and the proportion of identification can also vary significantly. The selection of the workflow may depend on the mentioned options. For this study, both the workflows were tested and the workflow yielding better results is presented. For future studies, it is recommended to consider the mentioned observations before workflow selection.

2.6. Conclusions

Using cultures of *E. coli*, *B. subtilis*, and *S. cerevisiae*, I have shown that it is possible to use the ChipFilter device to prepare microbial samples for proteomic analysis.

For this purpose, I introduce modifications concerning the lysis buffer, sample loading conditions, and elution of peptides. It allowed also to reduce of cross-contamination between the samples. Different experimental workflow has been introduced and will be further discussed.

The results, in terms of the number of proteins and peptides identified, are in line with those described in the literature.

3. ChipFilter sample preparation performance assessment with existing workflows

3.1. Introduction

We have seen in the previous Chapter that it is possible to analyse microbial cells using the ChipFilter sample processing. The results were in line with those of the literature. However, to assess the value of the ChipFilter, it is essential to compare its performance with other sample preparation protocols used in the laboratory. This is the purpose of this chapter.

To reach this goal, I compared the performance difference between the ChipFilter workflow from the existing workflows including FASP, in-gel and in-solution proteolysis. I compare the workflows for the sample preparation of mixed cells in terms of time taken for each preparation, identification of the proteins or peptides, distribution of proteins between the three species, handling efficiency with low cell number samples, and peptides properties including length, the theoretical mass of the positive ions, missed cleavages, hydrophobicity, and PSMs generated. Finally, the repeatability of the ChipFilter workflow was compared to other workflows and between different cell numbers for mixed cells. The entire results of this chapter and a portion of results from chapter 5 have been published as an article (<https://doi.org/10.1101/2023.01.18.524548>) attached in the annex.

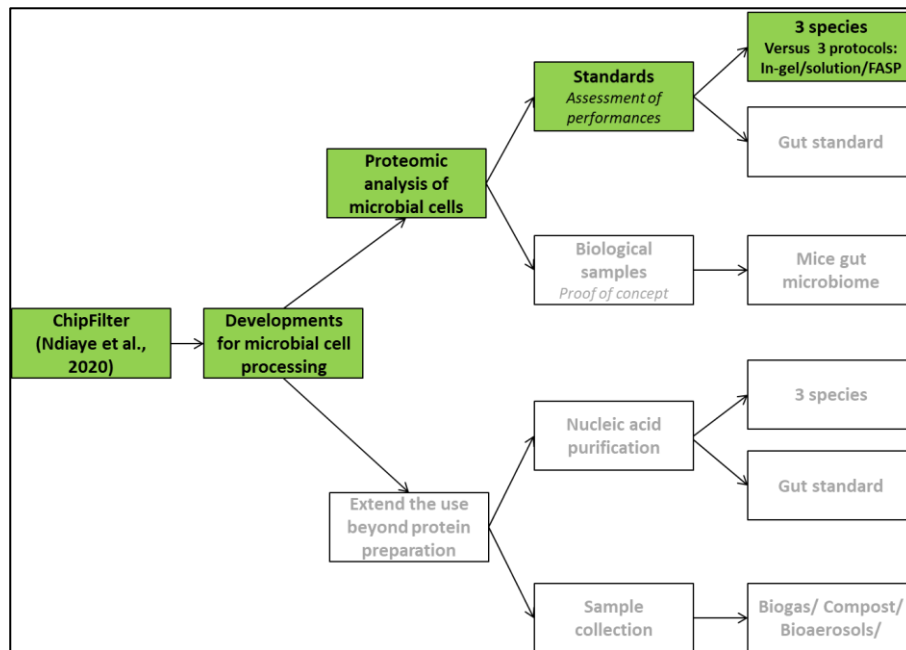


Figure 35: Steps discussed in chapter 3

3.2. Material and methods

3.2.1. Biological material and chemicals

Octyl- β -D-glucopyranoside (ODG), protease inhibitor, dithiothreitol (DTT), ammonium bicarbonate (ABC), and iodoacetamide (IAA) were purchased from Sigma-Aldrich. Trifluoroacetic acid was purchased from Thermo Fisher Scientific and acetonitrile (ACN) from Fisher Scientific. Trypsin Gold, Mass Spectrometry Grade was from Promega. Lysozyme (50 ng/ml) was from Thermo Fisher Scientific. Luria-Bertani (LB) agar (Merck) plates were inoculated with *Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacillus subtilis* and incubated at 37°C overnight. Inoculum into LB broth was made and cells were cultivated until the optical density reached 1 at 600 nm. Before harvesting, cells were counted using a glass slide under a microscope. The cells were collected by centrifugation at 2400 g for 5 minutes at room temperature. Single wash with phosphate buffer saline was performed before pelleting and stored at -80 °C until further use. Samples are named mix 1, mix 2, and mix 3 based on the total number of cells that is present. The mixed cells comprised three species *E. coli* (Gram-negative bacteria), *B. subtilis* (Gram-positive bacteria), and *S. cerevisiae* (fungi). The cells were grown as mentioned in section 2.4 and mixed in definite cell numbers of 1×10^6 cells/species for mix 1, 1×10^4 cells/ species for mix 2, and 1×10^2 cells/ species for mix 3. The cell number was achieved for each species separately by serial dilution and then mixing them with the appropriate cell number. This is a uniform mix of all three species at their respective cell number.

3.2.2. Methods

Four procedures have been used (figure 36).

ChipFilter procedure-

The ChipFilter WF1 (see part 2.2.3) with cellular injection using a piston syringe and the pump was done.

Modified Filter-aided Sample Preparation (mFASP) procedure-

This protocol was modified from the original FASP protocol (Wiśniewski et al., 2009) to imitate the conditions used in the device. This protocol is used to compare the two technologies that have a very similar designs under identical chemical conditions.

According to this protocol, the sample was resuspended in Lysis buffer I (sample/Lysis buffer I 1:5 v/v) and incubated at 37 °C for 15 minutes. The contents are transferred to Microcon-10 kDa Centrifugal Filter Units (Merck) and centrifuged at 15 000 g for 30 minutes with the flowthrough removed. The reduction was done with 20 mM DTT at 37 °C for 2 hours and after the tube was centrifuged at 15 000 g for 30 minutes with

the flowthrough removed. The reduction was performed with 50 mM IAM for 2 hours in dark at 37 °C and after the tube was centrifuged at 15 000 g for 30 minutes with the flowthrough removed. Two washes with 50 mM ABC buffer were performed to remove the reagents. Digestion was performed with Trypsin at a final concentration of 2 µg in a 50mM ABC buffer for 20 hours at 37 °C. Elution of peptides was done by centrifugation at 15 000 g for 20 minutes. Elution was done twice with 50 mM ABC to ensure maximum peptide recovery. Acidification of the peptides was performed with Trifluoroacetic acid (TFA) (Thermo Fisher Scientific) at a final concentration of 0.1% (v/v). Desalting was performed with C₁₈ zip tips (Merck) as per manufacturer guidelines. Peptides were dried after desalting and stored at -20 °C until injection to LC-MS. Before LC-MS/MS analysis the dried peptides were resuspended in 7 µl of 0.1% TFA and a 6 µl injection was made.

As we know already that ChipFilter is the miniaturized structure of the traditional filter tubes used in FASP protocol, I have selected to take the mFASP protocol that closely resembles in operation (no pretreatment, same lysis buffer) and reagents used to the ChipFilter workflow with exception only for proteolysis time and peptide loading into the LC-MS/MS.

In-gel procedure-

The sample was resuspended in 1X Laemmli buffer with β-Mercaptoethanol (Bio-Rad). The sample was heated at 95 °C for 5 minutes to improve cell lysis. Next, 40 µl lysate was loaded into 12% precast gels in a Tris-Glycine-SDS running buffer at a voltage of 80V for 15 minutes. A short SDS-PAGE migration was used to restrict the proteome to a short band before separation (Hartmann et al., 2014). The gel was stained with Coomassie brilliant blue R-250 and the protein band was sliced and washed with excess double distilled water under agitation (700 rpm/min) at room temperature. The gel was dehydrated with 100% acetonitrile (ACN, Fisher Scientific) and dried using a Speedvac system (Thermo Fisher Scientific). The reduction was performed with DTT (10mM) in 50mM ABC buffer at 56°C with agitation for 30 minutes. The excess solution was removed and the gel was dehydrated again with 100% ACN. Alkylation was performed next with IAA (55mM) in a 50mM ABC buffer with agitation at 37°C for 20 minutes in the dark. The excess solution was removed and replaced by a 50 mM ABC buffer. Dehydration was performed with 50% and 100% ACN and completely dried by vacuum drying. 200 µl of trypsin solution at the concentration of 12.5 ng/µl was prepared with 50mM ABC buffer and rehydration of the gel was done for 45 minutes at 4 °C. Next, the excess solution was removed and digestion was performed for 20 hours at 37 °C with agitation. Extraction of the peptides was performed in two steps acidification and dehydration of the gel. Acidification of the gel was performed by the addition of 100ul of 1% TFA and 5% of formic acid separately for 15 minutes under agitation at 37 °C. The solution was transferred to a

new tube. Again, the gel was dehydrated with 100% ACN for 15 minutes under agitation at 37 °C and the solution was transferred to the elution tube. The elute was dried by vacuum drying and resuspended in 20 µl of 1% (v/v) formic acid and sonicated for 5 minutes at 37 °C. Desalting was performed with C₁₈ zip tips as per manufacturer guidelines. Peptides were then dried and stored at -20 °C until injection into LC-MS. Before LC-MS/MS analysis the dried peptides were resuspended in 7 µl of 0.1% TFA and a 6 µl injection was made.

The in-gel workflow was taken to compare the ChipFilter work to a gold standard workflow that is well-known in metaproteomics. By using this workflow, the performance of the ChipFilter as well as the caveats in the existing workflows can be better understood.

In-solution procedure –

The sample was resuspended in Lysis buffer II (6M Urea, Tris 150mM, Protease Inhibitor 1X, 1% (w/v) octyl-beta-D-glucopyranoside) with sample/Lysis buffer II ratio of 1:5 (v/v). Samples were subjected to mechanical disruption with glass beads of 425-600 µm (Sigma-Aldrich) in the ratio of 1:1 (v/v). Five cycles consisting of 30s vortex and 60s in ice were repeated as bead-beating pre-treatment. The lysate was carefully removed from the beads and transferred to a new tube, where the reduction was performed with 20 mM DTT at 37 °C for 2 hours. In the same tube, alkylation was performed with 100 mM IAA at 37 °C for 2 hours in the dark. Precipitation of the proteins was performed with 10% TCA solution for 30 minutes at 4 °C, followed by centrifugation at 17 500 g at 4 °C for 1 hour. The protein pellet was washed with 80% and then 100% (v/v) ice-cold acetone, followed by drying using speed-vac. The protein pellet was solubilized in a 50 mM ABC buffer and digestion was performed with 2 µg of trypsin for 20 hours at 37 °C. Acidification of the peptides was performed with 0.1% (v/v) TFA. Desalting was performed with C₁₈ zip tips as per manufacturer guidelines. Peptides were dried after desalting and stored at -20 °C until injection to LC-MS. Before LC-MS/MS analysis the dried peptides were resuspended in 7 µl of 0.1% TFA and a 6 µl injection was made.

In-solution workflow offers the comparison between the requirement of pre-treatment (bead-beating) and tougher cell lysis buffer as this utilizes 6M Urea solution in addition to the lysis buffer used in ChipFilter workflow.

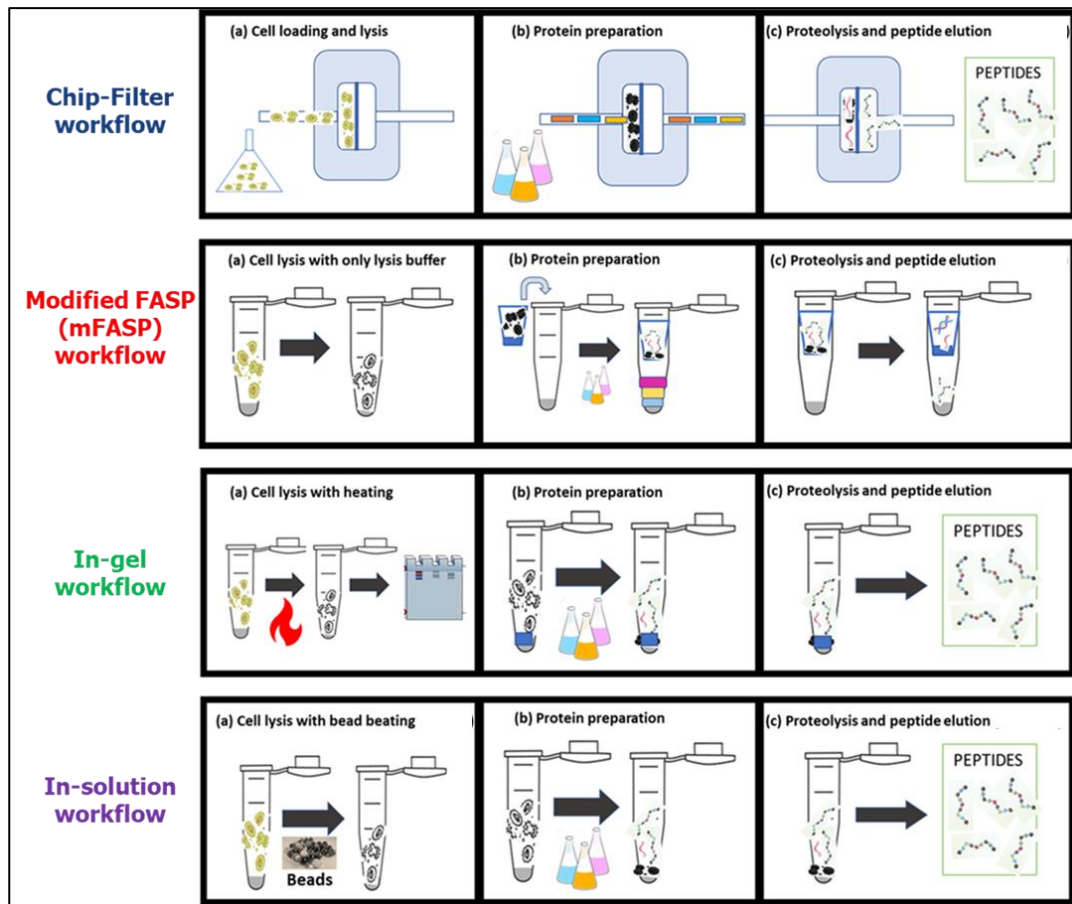


Figure 36: Steps used in different procedures discussed for microbial sample preparation for LC-MS/MS proteomics

3.2.3. LC-MS/MS analysis and protein identification

Samples were analysed by nanoLC-MS/MS in data-dependent acquisition (DDA) high-energy c-trap dissociation (HCD) mode using an RSLCnano UltiMate™ 3000 System coupled to a nanoESI Q-Exactive or Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). For the ChipFilter method, the peptides recovered in the trap column were separated on a capillary reverse-phase C18 column Pepmap 75 μm i.d. \times 50 cm length (ThermoFisher Scientific) at 45°C with a linear 120 minutes gradient elution from 2.5% to 60% of buffer B (water/acetonitrile/formic acid 10%:90%: 0.1% (v/v/v)) in buffer A (water/acetonitrile/formic acid 98%: 2%: 0.1% (v/v/v)) at a fixed flow rate of 220 nL/min. For mFASP, in-gel and in-solution methods, the dried peptides were resuspended in 7 μl of 0.1 % TFA solution (v/v) and 6 μl was used for single-shot injection. Trapping was done with C18 Pepmap 300 μm i.d. \times 5 mm length column (Thermo Fisher Scientific) and analysed in nanoLC MS/MS with a 120 minutes gradient as described earlier. Mix-1 and Mix-3 samples analysis were performed with a Q-Exactive mass spectrometer operated in nanoESI at 1.7 kV. Full MS survey scans were recorded over the m/z range of 400 – 2000

with a resolution of 70,000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 100 ms. Up to 15 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 30, with a precursor isolation window at 2 m/z, resolution set at 17,500 with AGC value at 1E5 with a maximum injection time of 120 milliseconds. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 seconds.

3.2.4. Data processing

Spectra were processed using Proteome Discoverer (PD) v2.4 (Thermo Fisher Scientific). For mixed cells, the Mascot search engine (Matrix Science Mascot 2.2.04) was used against the all-taxonomy SwissProt database (release 2022_03: 568002 sequences; 205171419 residues). The database search was performed with the following parameters: MS and MS/MS mass tolerance 10 ppm and 0.02 Da respectively, trypsin specificity with up to 2 missed cleavages, partial Carbamidomethylation (C), Deamidation (NQ) and Oxidation (M). Proteins with at least one high-confidence peptide and six amino acids were validated. Target FDR was set at 0.01. To perform label-free quantification in PD, the Minora feature detection tool was used in the processing workflow and the precursor ions quantifier in the consensus workflow with consideration for both unique and razor peptides. The abundance value obtained was used to make the scatter plot and determine the correlation using tools in R (v 4.0.3). The Kyte–Doolittle hydrophobicity index has been calculated from a published package in R-Journal (Osorio et al., 2015). All the experiments were performed in four replicates for mixed cells. Statistical tests to identify significance (unpaired t-test) were performed and values shown always represent mean ± standard error of the mean (SEM). The mass spectrometry data have been deposited on the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) [px-submission #637484].

3.3. Experimental time

Before looking at the results of the four types of proteolysis, it is possible to consider the time taken for each procedure (figure 37).

The shortest time taken was 377 minutes by ChipFilter workflow. The reason is the usage of shorter proteolysis duration. As all the workflows except ChipFilter require overnight digestion, whereas only 120 minutes was necessary for ChipFilter workflow. A shorter digestion time in the ChipFilter workflow is preferred as the peptides start to exit the chamber after two hours and it was necessary to stop the digestion at a shorter time if the flow of liquid is to be used. In continuous flow digestion, reported in this work, the digestion time has to be shorter to avoid the loss of peptides. For reactions like cell lysis, reduction, alkylation, and washing the total time taken was shorter for in-gel workflow

(190 minutes) as compared to ChipFilter workflow (227 minutes). The cell lysis step for the ChipFilter was 4 times more than the mFASP workflow and it takes the second most duration after proteolysis for the ChipFilter workflow. As there is no pre-treatment step in the ChipFilter workflow, it is critical to increase the cell lysis by passing the lysis buffer for a longer duration. An important point is that only the ChipFilter workflow is completely automated.

However, unlike other workflows that allow multiplexing, the ChipFilter workflow is currently being used with one sample at a time basis and not by multiplexing as the instrumentation (M-switch valve) has not been modified to accommodate multiple outlets. The development to allow multiplexing is possible for ChipFilter too without changing the design of the device, but by only adding external elements.

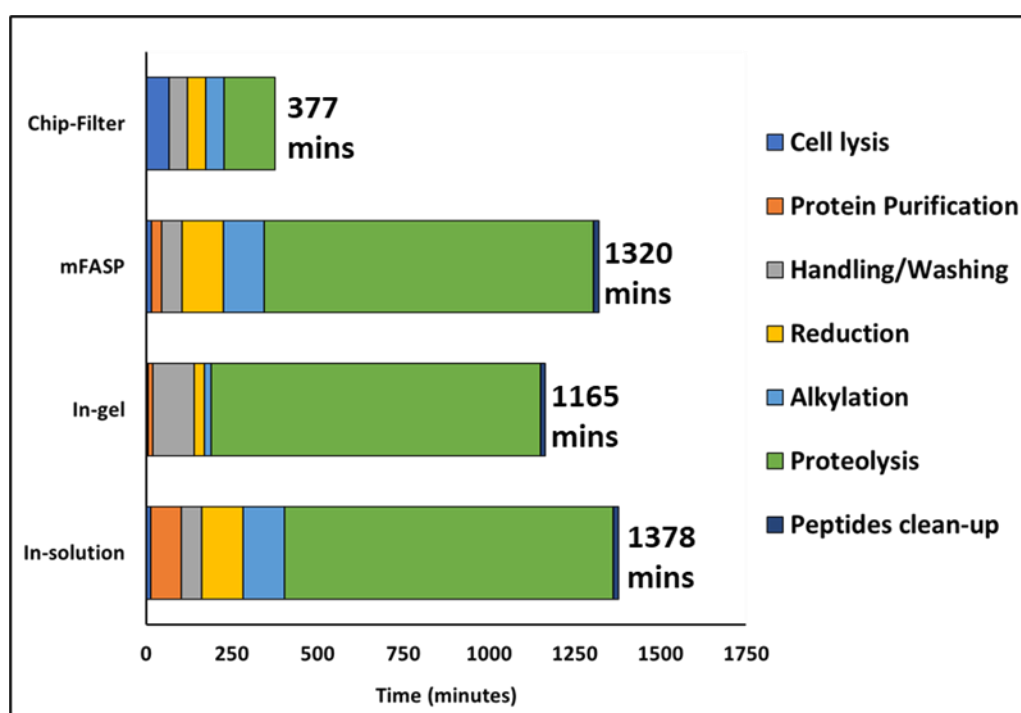


Figure 37: Time taken for seven key steps in proteomics sample preparation by different procedures discussed

3.4. Protein identification and distribution

The performance of the protocols was evaluated by considering the number of target proteins and total peptides identified. For the concentration of 1E6 cells, the ChipFilter method achieved a mean protein identification of $1\,999 \pm 54$ whereas mFASP reached $1\,818 \pm 91$, in-gel $1\,504 \pm 22$, and in-solution $1\,665 \pm 15$ (figure 38A). The ChipFilter workflow has significantly higher identification than in-gel (p-value 0.002) or in-solution procedure (p-value 0.013) and is as competent as mFASP (not statistically significant) for

mix 1. Similarly, at the peptide level, a significantly higher identification number is obtained for the ChipFilter method than other methods as shown in figure 38 B. $9\,770 \pm 108$ peptides were identified using the ChipFilter method, $8\,197 \pm 435$ with mFASP, $8\,218 \pm 262$ with in-gel, and $7\,805 \pm 178$ with in-solution. A significantly higher identification is seen for ChipFilter workflow than other workflows (p-value less than 0.05).

Efficient catalysis in confined spaces and reduced loss of materials can be attributed to the success in higher identification for the ChipFilter procedure. Further, the distribution of the proteins and peptides among the procedures was studied. Due to the specificities of each protocol, it is observed that proteins identified by the four procedures are not the same, in other words, there are still proteins identified exclusively by one protocol, and proteins identified from different protocols. At the protein level (figure 38C), 17.5% (598 proteins) of the total proteins were identified exclusively by the ChipFilter method, whereas the other methods had less than 5.1% of specific proteins. The protein identification common between all the methods was about 38.5% (1 319 proteins). When we consider proteins identified exclusively by one protocol, we observe that a three-fold increase in the identification of exclusive proteins was achieved by the ChipFilter procedure. The superior identification of the proteins only by ChipFilter workflow suggests that it can be used suitably for metaproteomics sample preparation.

At the peptide level (figure 38D), 32% (8 395 peptides) of the total peptides were identified only in the ChipFilter method. Interestingly, the peptide common to all the methods is only 8.8% (2 309 peptides) of the total population, indicating that each method characteristically generates different peptide fractions. The peptide distribution suggestively varies as compared to the protein identification for each protocol for the ChipFilter workflow.

Another crucial parameter to be validated is the representation of all the species in a community for their protein identification. It is well known that the microbiome will not necessarily have the same cell number throughout and the difference in cell number between the participating species is a characteristic trait. Furthermore, the composition of any microbiome will include a variety of species belonging to bacteria, protozoa, viruses, archaea, or eukaryotes (fungi and algae). For a good sample preparation procedure, it is imperative to enable the processing of most of the species to increase identification coverage. The workflow must be capable of handling low cell number populations and as well lyse/recover proteins from different species having structurally complex cell characteristics. The cells used in mixed cells were selected based on an earlier report (Hayoun et al., 2019) and on the basis that the proteomic/genomic characterization of the three species is fully done with reference databases completely

available. This minimizes the complexity arising due to MS data treatment. Further, the three species represent the three domains of Gram-negative bacteria, Gram-positive bacteria, and fungi that are mostly found in microbiome communities. Also, the purpose of using different cell numbers of the population helps to understand the commonality or differences created by the ChipFilter workflow during sample processing.

A universal sample preparation workflow should facilitate the efficient lysis of different types of cells and extract the proteins from different cellular fractions. A comparison of the protein origin from three microbial cells (figure 38E and table 3) indicates that the ChipFilter identifies the highest percentage for Gram-positive bacteria *B. subtilis* (12%) and Gram-negative bacteria *E. coli* (37%) in comparison to other methods, whereas the highest percentage of fungi *S. cerevisiae* was recorded for in-solution (69%) method. Increased fungal proteins in the in-solution method can be reasoned due to the introduction of a pre-treatment step which was lacking in other methods. The in-gel method has poor identification for Gram-positive bacteria. The usage of SDS containing Laemmli buffer can be the reason as the peptidoglycan cell wall of Gram-positive bacteria was not efficiently denatured by SDS (Mahalanabis et al., 2009). The ChipFilter method can identify all three cell types considerably better than other methods without pre-treatment.

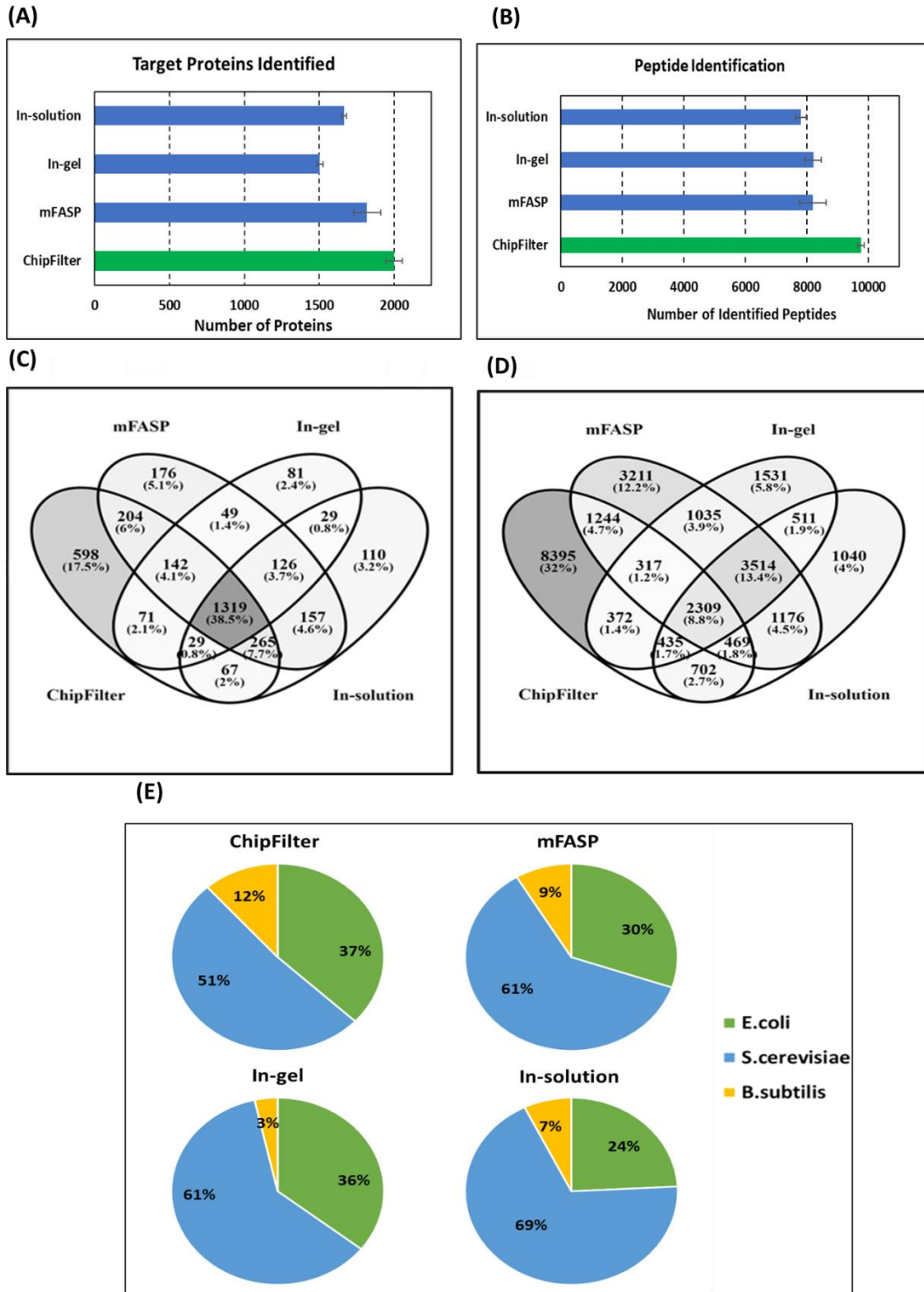


Figure 38: Performance assessment of ChipFilter workflow compared to other procedures for Mix 1. (A) Number of identified proteins and (B) peptides. Distribution of the (C) proteins and (D) peptides identified. (E) Distribution of the protein identified according to the species

Table 3: Number of proteins identified for each microbial species in mix 1 with every procedure

	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>B. subtilis</i>	Total
ChipFilter	738 ± 60	1031 ± 101	230 ± 24	1999 ± 54
mFASP	548 ± 18	1113 ± 63	157 ± 10	1818 ± 91
In-gel	536 ± 14	916 ± 10	53 ± 8	1504 ± 22
In-solution	403 ± 3	1141 ± 17	121 ± 2	1665 ± 15

3.5. Low cell number processing

The robustness of the ChipFilter workflow was tested for low cell number samples (10E2 of each species) to assess the identification and distribution of peptides and proteins. The same parameters have been studied as those of the previous section.

The sensitivity of the ChipFilter method was tested for low cell number samples (Mix 2) to assess the identification and distribution of peptides and proteins like in mix 1. The mean number of target proteins and total peptides identified are 163 ± 18 and 1162 ± 126 respectively (figure 39A and 4B). The number of protein identifications is similar for the ChipFilter method as compared to the in-solution method (172 ± 28 proteins), but the number of peptide identification is significantly higher than mFASP (p-value 0.015) and in-gel (p-value 0.016). This suggests that ChipFilter workflow is adapted at this cell density.

The distribution of the proteins (figure 39C) indicates that the percentage of proteins exclusively identified by the ChipFilter method (20.6%, 79 proteins) and in-solution method (17.5%, 67 proteins) are very close, unlike in mix 1. Furthermore, the percentage of proteins identified by both methods is about 45.9%, suggesting that the protein representatively is very similar. Of the total proteins identified, only 27.9% were not identified by the ChipFilter method. For the peptide identifications shown in figure 39D, the peptides exclusively identified by the ChipFilter method is 45.1% (1270 peptides). It is roughly three times higher than the peptides exclusively identified by the in-solution method (13.4%, 378 peptides). The proportion of peptides not identified by the ChipFilter method is 30.4% (858 peptides). As observed with Mix 1, the distribution of peptides or proteins is not similar, indicating that sample preparation methods handle low cell density samples differently. Finally, the representation of protein identifications for different species with the methods is indicated in figure 39E and table 4. Like in Mix 1, the distribution of the proteins identified is represented for all three species for the ChipFilter method. In the case of the identification of *B. subtilis*, the ChipFilter identifies 9% of total proteins whereas, it is less than 4% for the other methods. Furthermore, the protein identification for *S. cerevisiae* for the in-solution method is 87% of the total proteins identified which is higher than the other methods which have about 50% to 60%

of this species. The ChipFilter method is a universal sample preparation method for simultaneously identifying proteins in a single analysis even at low cell numbers.

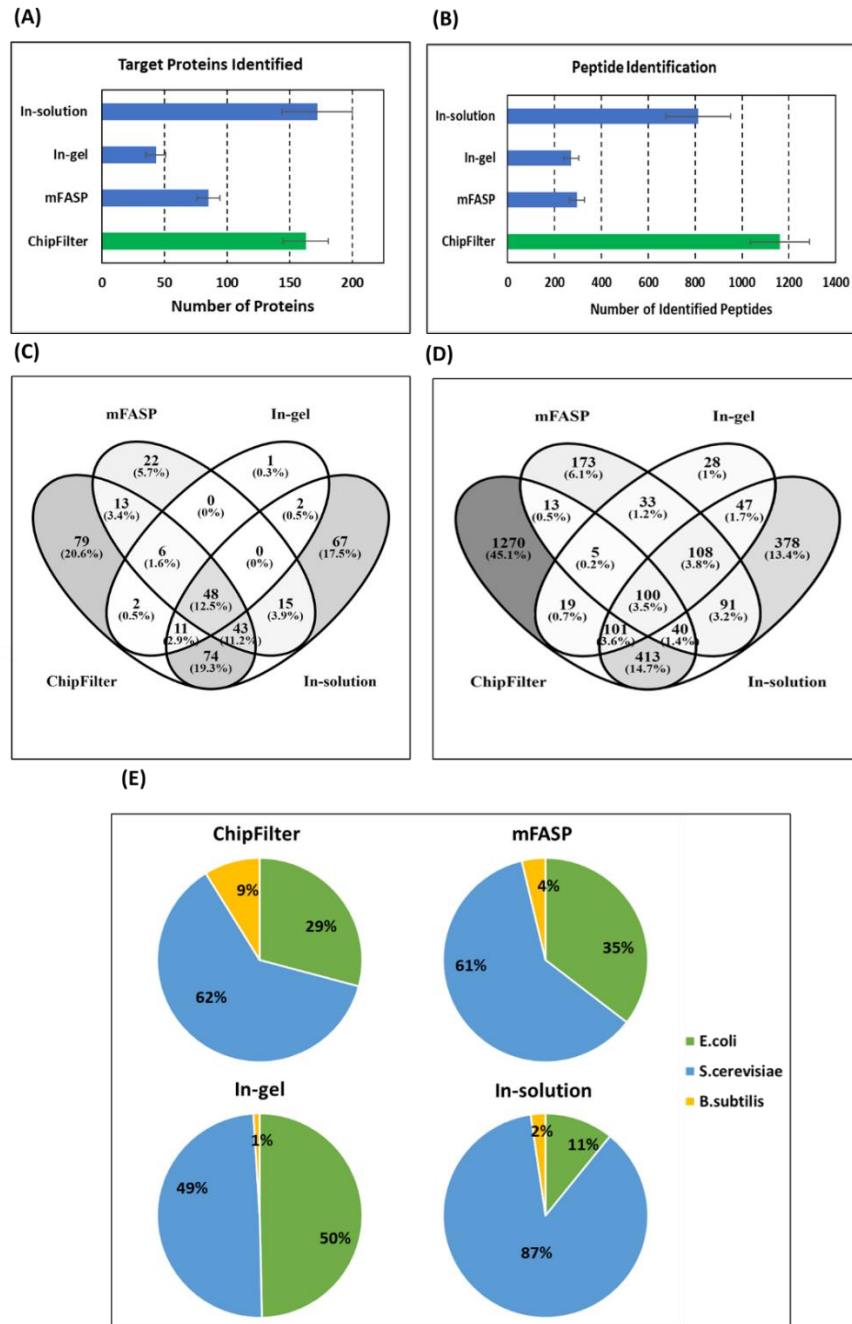


Figure 39: Performance assessment of ChipFilter workflow compared to other procedures for Mix 3. (A) Number of identified proteins and (B) peptides. Distribution of the (C) proteins and (D) peptides identified. (E) Distribution of the protein identified according to the species.

Table 4: Number of proteins identified for each microbial species in mix 3 with every procedure

	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>B. subtilis</i>	Total
ChipFilter	46 ± 3	104 ± 21	13 ± 5	163 ± 18
mFASP	30 ± 4	52 ± 6	3 ± 1	85 ± 9
In-gel	21 ± 3	22 ± 5	1 ± 0	43 ± 8
In-solution	18 ± 3	149 ± 24	4 ± 1	172 ± 28

3.6. Peptide level variations

Comparing the number of peptides or proteins is not sufficient to compare different protocols. It is also necessary to study the physio-chemical characteristics of these peptides, to identify possible identification biases. In that regard, several parameters include the number of positive ion mass (MH⁺) values, amino acids length of the peptides, missed cleavages, peptide spectral matches (PSMs), and hydrophobicity index of the peptides.

The peptide mass and length have been studied to understand the nature of peptides' physical properties (figure 40A). In mix 1 for the ChipFilter protocol, the mean mass of peptides with less than 2000 Da is 32% (3137 ± 149 peptides), while for the mean mass of peptides with more than 2000 Da is 68% (6633 ± 109 peptides). In mix 1 for mFASP, in-gel and in-solution workflows, the average mass of peptides with less than 2000 Da is 73% (5879 ± 232 peptides), while that greater than 2000 Da is 27% (2194 ± 142 peptides). Due to the reduced concentration of proteins in mix 3, the enzymatic proteolysis is expected to be better than in mix 1. This is true for ChipFilter, as the number of peptides for ChipFilter workflow less than 2000 Da is 56% (1162 ± 73 peptides) has increased and that greater than 2000 Da was 44% (5077 ± 62 peptides) is reduced as compared to mix 1. Also, for other procedures, an increased percentage of smaller mass peptides was seen. For mFASP, in-gel and in-solution procedures, the average number of peptides less than 2000 Da is 89% (460 ± 67 peptides), and that greater than 2000 Da was 11% (65 ± 14 peptides). These results are in line with the number of missed cleavages. All mass range distributions remained significantly different when ChipFilter was compared with other procedures for mix 1. Whereas in mix 3 mass range between 1000 – 1500 Da was not significantly different for any method, and the mass range 1500–2000 Da was not significant between ChipFilter and in-solution procedures.

The size of the peptides has also been studied in terms of the number of amino acids (figure 40C). ChipFilter protocol resulted in the identification of 42% (4063 ± 162 peptides) of peptides shorter than 20 amino acids and 58% (5707 ± 83 peptides) larger than 20 amino acids. For mFASP, in-gel and in-solution workflows, peptide lengths less

than 20 amino acids are around 80% (6477 ± 241 peptides) and the peptides longer than 20 amino acids are 20% (1596 ± 116 peptides). Similarly, in mix 3 (low cell density), for ChipFilter the proportion of peptides shorter than 20 amino acids was 66% (767 ± 80 peptides), whereas that longer than 20 amino acids were 34% (395 ± 53 peptides). For mFASP, in-gel and in-solution workflows, peptide lengths less than 20 amino acids are approximately 91% (410 ± 59 peptides), and the peptides longer than 20 amino acids are only 9% (50 ± 11 peptides). As seen earlier for MH+ values, reduction in protein density reduced the length of the peptides due to better proteolysis and ChipFilter workflow has significantly longer peptides than other workflows compared indicating that it can generate novel peptides fragmentation spectra that may allow better identification of protein sequences. All peptide size range distributions remained significantly different when ChipFilter was compared with other procedures for mix 1. Whereas in mix 3 size range between 1 – 10 amino acids were not significantly different between ChipFilter and mFASP procedures, and the size range of 10–20 amino acids was not significant between ChipFilter and in-solution procedures.

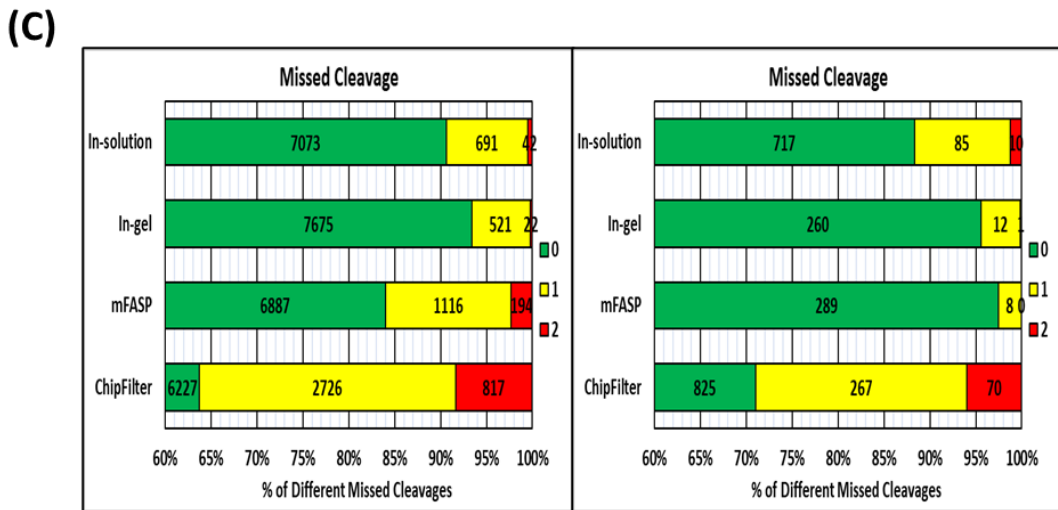
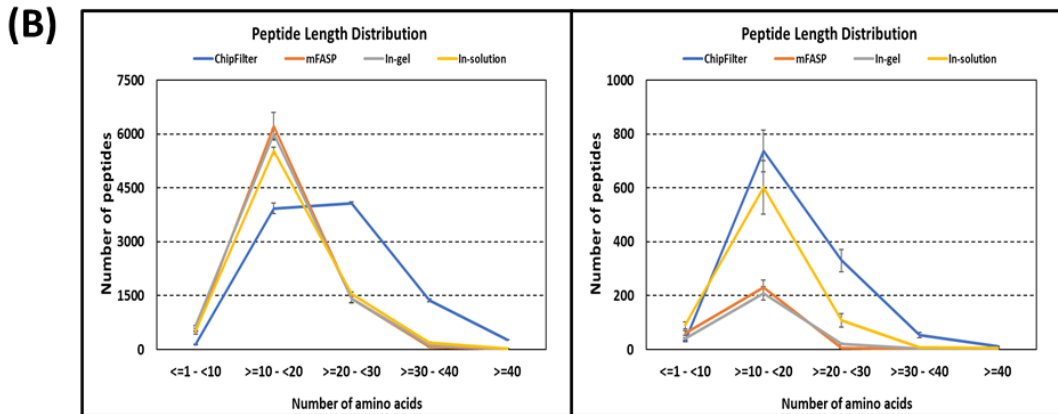
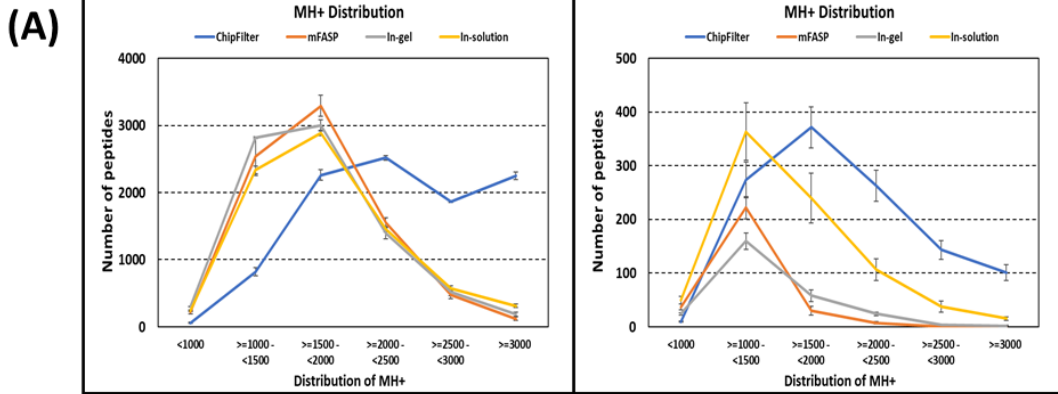
In figure 40C, the percentage of missed cleavages is indicated for mix 1 and mix 3 for the peptides identified with the four protocols. ChipFilter method has the highest percentages of 1 and 2 missed cleavages with mix 1 (28% and 8% respectively) and mix 3 (23% and 6% respectively). The decrease in the number of missed cleavages when the number of cells decrease suggests better proteolysis for low cell number samples in microfluidic systems, which may be attributed to a higher enzyme/substrate ratio. The percentage is also considerably higher in the mFASP procedure for mix 1 (14% and 2% respectively) than in-gel or in-solution, but for mix 3 it is lower (3% and 0% respectively). In-solution method is consistent between the two cell numbers with around 9% for 1 missed cleavage and 1% for 2 missed cleavages. The in-gel method has the least. This could be because peptides containing missed cleavage are longer than their counterpart without missed cleavage and would have a harder time being eluted from the gel. A high percentage of missed cleavage in the ChipFilter method can be attributed to several factors, crowding of the proteins within a confined space limiting entropy, shorter proteolysis duration, and insufficient transfer of the enzyme on the membrane trapping the proteins. Although, external circulation of the buffer at a low flow rate is applied to increase enzyme-substrate interaction, the possibility of peptides exiting the system with incomplete digestion is higher. That is why I proposed an alternative to workflow 1, named workflow 3, wherein the circulation of buffer during proteolysis was avoided to ensure the exit of partially digested peptides and therefore decrease the percentage of missed cleavage. Although the results from both workflows were not compared to evaluate the percentage of missed cleavages, I would like to propose this possibility for future experiments that require complete proteolysis. In the context of shotgun

metaproteomics where it is critical to identify more peptides, having missed cleavage proves to be very advantageous as in this case ChipFilter workflow identifies more proteins than other methods.

PSMs are useful to identify the sequence of the peptide's spectra produced in tandem with MS by assigning scores to the theoretical spectra that can be generated for a specific peptide fragment *in-silico*. This method depends on the fragmentation pattern responsible for the generation of the peptides to compute the scores. Often it is considered that the generation of more PSMs per peptide is necessary. In this regard, the PSMs distribution was studied (figure 40D). In contrast to the other three protocols, for which the number of PSMs is quite similar, the ChipFilter shows a very large number of peptides identified with less than 10 PSMs. It should be noted that this is not at the expense of the number of peptides identified with more PSMs, whose values are close to those obtained with the other protocols. This explains why, although the proportion of peptides identified with less than 10 PSM is significant, the number of peptides identified is greater with the ChipFilter.

It is interesting to note that the higher percentage of missed cleavages in the ChipFilter workflow could have contributed to the generation of peptides with lower PSMs. As the PSMs are identified computationally, the generation and determination of many decoy peptides that match with missed cleavages will be unlikely.

Finally, the hydrophobicity of the peptides generated was studied using the Kyte-Doolittle hydrophobicity index (Kyte & Doolittle, 1982) (figure 40E). Most of the peptides were accumulated in the range between -1 to +1 which is indicative of a neutral mixture with slight hydrophobic or slight hydrophilic peptides. This trend was observed for all the workflows including ChipFilter for both the cell mixtures. Although the number of peptides identified for every range was significantly different between the procedures, the distribution remained the same. One interesting observation that was made was the decrease in the mean number of hydrophilic peptides (h-index ≥ -1) for ChipFilter (269 ± 20) compared to other procedures (589 ± 17) in mix 1. This was not the case for mix 3.



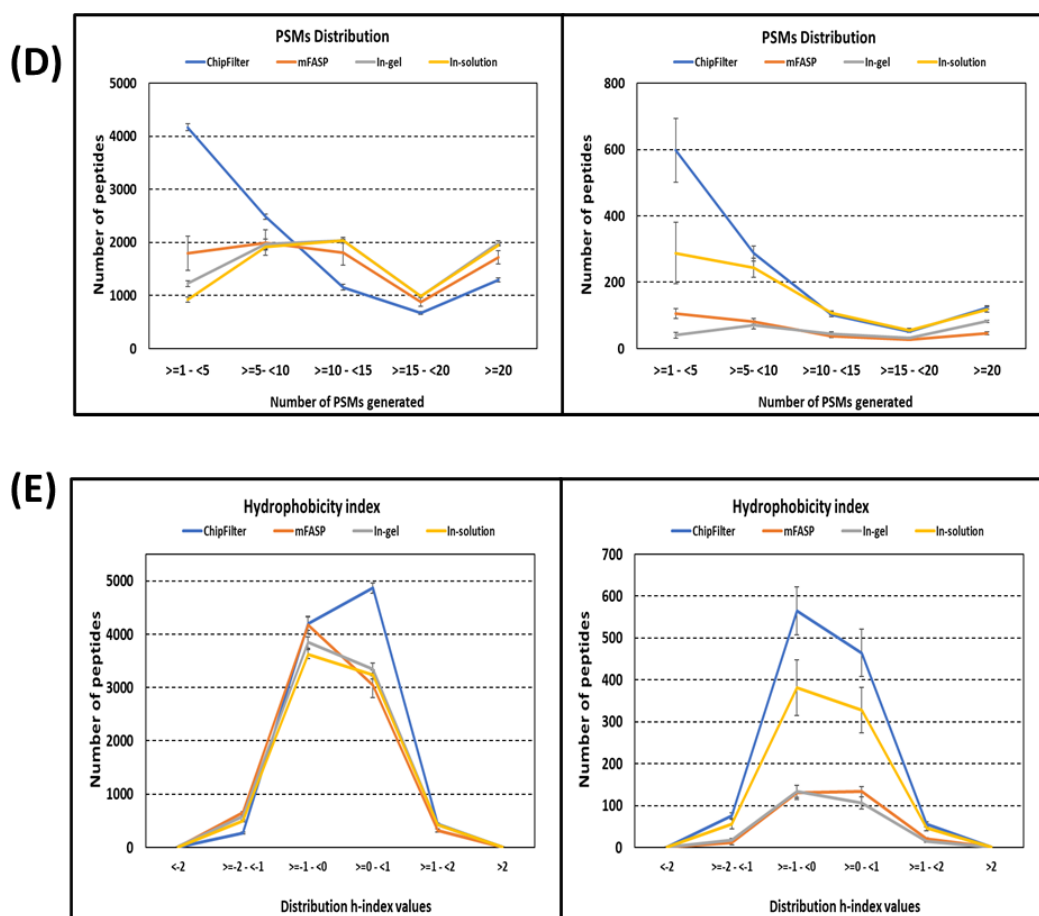


Figure 40: Physical chemical characteristics of the peptides generated by different preparation procedures for mix 1 (left) and mix 3 (right). (A) Positive ion mass (MH⁺), (B) amino acid length distribution, (C) percentage of missed cleavages, (D) peptide spectral matches (PSMs) distribution, and (E) Kyte-Doolittle hydrophobicity index

3.7. Repeatability

Repeatability is often difficult to achieve in metaproteomics due to the variations in biomass between samples but is critical. To assess the repeatability of the ChipFilter procedure, the correlation of the identifications for the LFQ of the identifications of 4 replicates was determined for mix 1 and 3 (Figure 41A and 41B respectively)

Correlation is shown to be in the positive direction, indicating that changing the parameters in one replicate induces similar changes in other replicates. The correlation coefficient values for ChipFilter workflow are in the range of 0.76 to 0.83 for mix 1, indicating a very high positive correlation between the replicates for high cell numbers. Whereas, the correlation coefficient values for ChipFilter workflow are in the range of 0.59 to 0.69 for mix 3. The reduction in the correlation can be due to the reduced intensities obtained at lower cell numbers. Even though the peptide identifications are

significantly high for ChipFilter workflow in mix 3, the correlation between the replicates is considerably small.

The comparison of the correlation coefficients for every procedure indicates that the ChipFilter workflow has high variability among the different values. The mFASP (0.69 to 0.92 for mix 1 and 0.75 to 0.89 for mix 3), in-solution (0.84 to 0.87 for mix 1 and 0.76 to 0.83 for mix 3), and in-gel (0.86 to 0.87 for mix 1 and 0.75 to 0.80 for mix 3) procedures have the shorter range as compared to ChipFilter workflows. Scatter plots and correlation values can be seen in the annex.

In conclusion, ChipFilter workflow generates replicates that have a positive correlation in mix 1 and 3 indicating that reproducibility is satisfactory. In my opinion, further improvements in proteolysis conditions can be made to increase the reproducibility of ChipFilter workflows.

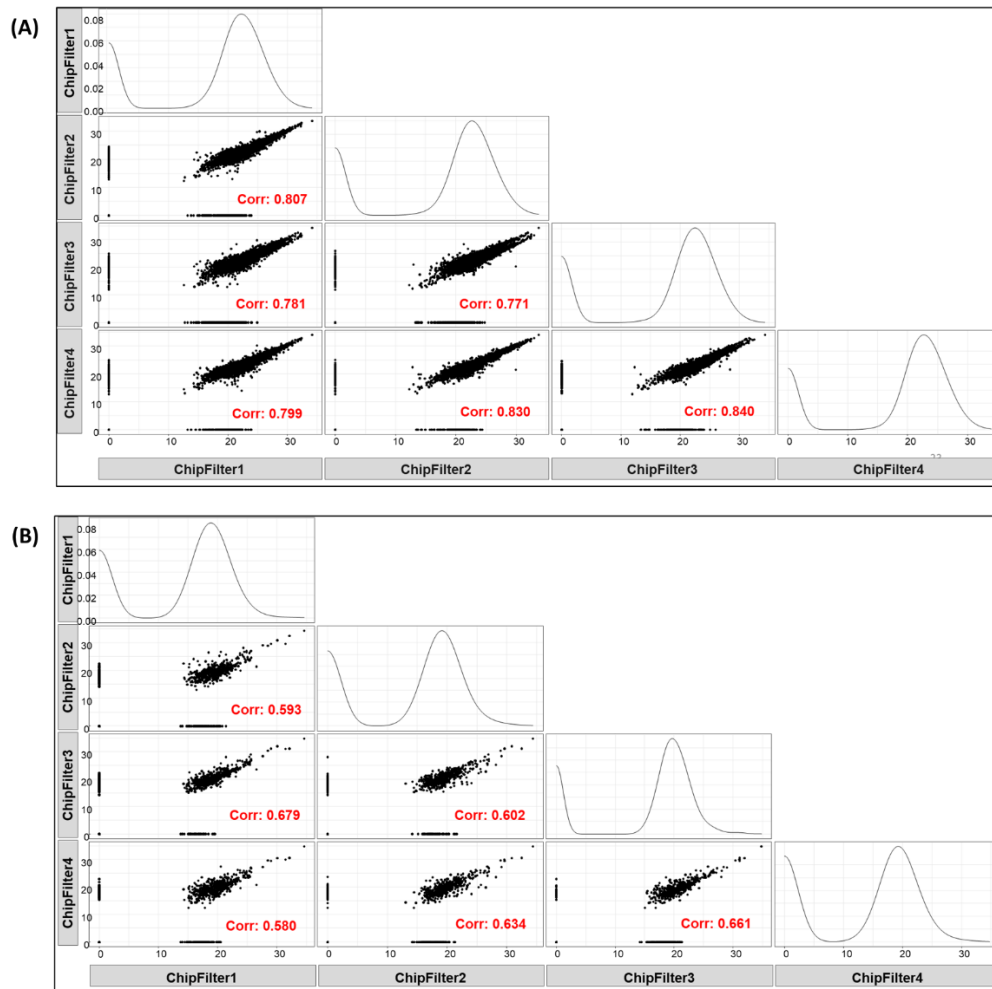


Figure 41: Label-free quantification (LFQ) intensity for ChipFilter workflow for mix 1 (A) and mix 3 (B)

3.8. Conclusions

We have proved, in Chapter 2, that it is possible to perform microbial cell sample preparation in the ChipFilter device. The next step was to assess the efficiency of this preparation. To do that, we compared the results of two mixtures of *E. coli*, *B. subtilis*, and *S. cerevisiae* at two cellular concentrations (10E6 and 10E2), after preparation in the ChipFilter with three standards protocols: in-gel, in-solution, and mFASP.

ChipFilter was the most efficient procedure in terms of protein and peptide identifications. In addition to better identification, the distribution of the proteins and peptides among the different protocols indicated ChipFilter to have a higher proportion of exclusive peptides over others.

Probing the physiochemical properties of the peptides generated by different workflows indicates the ChipFilter method has a higher percentage of missed cleavages that results in peptides with longer-length amino acids and higher MH+ values than other workflows. In terms of the PSMs generated, the ChipFilter workflow had a higher portion of peptides with less than 10 PSMs while other workflows had more peptides with higher than 10 PSMs. In terms of the hydrophobicity index, there was no observed difference between the different protocols.

The estimation of the repeatability between different replicates using the LFQ method showed good repeatability for ChipFilter workflow but it was lower than that of the other procedures.

As compared to mFASP, in-gel and in-solution protocols, the ChipFilter can offer operational advantages like complete automation and cell lysis without pre-treatment during sample preparation.

Before applying this protocol to the study of more complex samples, we explored the possibility of using the ChipFilter device for applications other than proteomic analysis.

4. Proteomic analysis of murine gut microbiome: a proof of concept

4.1. Introduction

Until now standard mixtures of cells were processed by ChipFilter workflow. These were simplified models, which allowed us to develop the conditions for the analysis of proteins and nucleic acids. However, there are many questions before using the ChipFilter for field sample study, for example, should the sample be pre-processed? Are there any inhibitors that should be removed? How to process the data? The purpose of this chapter is to establish the proof of concept for the use of the ChipFilter to study murine gut microbiomes.

Murine models have been useful for numerous studies. In that regard, the gut microbiome of the murine can serve as a model for human microbiome studies as the composition shared is 90% and 89% similar in phyla and genera, respectively (Krych et al., 2013). Murine models for gut microbiome studies offer several advantages the availability of a wide range of phenotypes and disease models, long-standing research information on murine genetics, and easier animal model to maintain. While the disadvantage is that it does not completely capture the human gut microbiome, different anatomy and genetic composition of the murine microbiome are dissimilar compared to the human microbiome (Nguyen et al., 2015).

Microorganisms isolated from the murine faeces, used as a representative of the murine gut microbiome, were processed using ChipFilter workflow 5. Their proteomic analysis was realized. Data analysis was performed by using two protein databases developed to identify the proteins from murine gut microbiome that lacks well-characterised protein databases. Finally, biologically relevant information was obtained for the murine gut microbiome metaproteomics (figure 42).

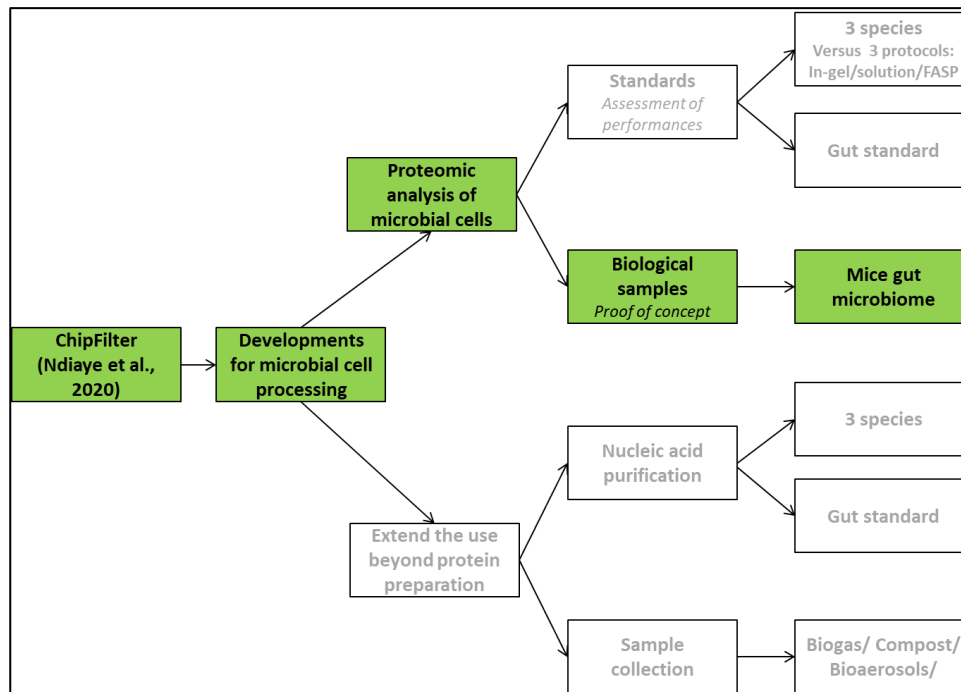


Figure 42: Steps discussed in chapter 4

4.2. Material and methods

4.2.1. Biological material

Murine faeces were collected from healthy female murine housed in Brain Plasticity Laboratory, ESPCI-PSL, Paris. Samples were collected, and aliquots of 25-30 mg were stored at -80 °C. All the experiments were done in three replicates.

4.2.2. Protein extraction

The samples were processed in four biological replicates. For the microbiome isolation from the collected faeces, the low-speed centrifugation method described by Wu *et al.*, (Wu *et al.*, 2016) was used. Briefly, 25 mg of faecal sample was suspended in 500 µl of PBS pH 7.4 with a 5% protease inhibitor cocktail. Physical disruption was done by vortexing and pipetting. Centrifugation at 250 g for 5 minutes at 4 °C was done and the supernatant was collected in a new tube. The pellet was resuspended and the same steps were repeated four times. Finally, the collected supernatant was centrifuged at 21 000 g for 30 minutes at 4 °C. The pellet was resuspended in 200 µl of lysis buffer comprising 1% (w/v) ODG, 150mM Tris-HCl pH 8.8, and protease inhibitor. The cell lysate was prepared by mechanical agitation with glass beads described in section 3.2. A final volume of approximately 350 µl was obtained. Protein estimation was done by the BCA method and

approximately 40 µg of protein lysate was loaded into the ChipFilter for sample preparation. Sample preparation in ChipFilter was performed by WF2.

4.2.3. LC-MS/MS analysis

The dried peptides were resuspended in 7 µl of 0.1 % TFA solution (v/v) and 2 µl was used for injection. Trapping was done with C18 Pepmap 300 µm i.d. × 5 mm length column (Thermo Fisher Scientific) and analysed in nanoLC MS/MS with a separated capillary reverse-phase C18 column Pepmap 75 µm i.d. × 50 cm length (ThermoFisher Scientific) at 45°C with a linear 120 minutes gradient elution from 2.5% to 60% of buffer B (water/acetonitrile/formic acid 10%:90%: 0.1% (v/v/v)) in buffer A (water/acetonitrile/formic acid 98%: 2%: 0.1% (v/v/v)) at a fixed flow rate of 220 nL/min. Peptides were analysed by nanoLC-MS/MS in data-dependent acquisition (DDA) high-energy c-trap dissociation (HCD) mode using an RSLCnano UltiMate™ 3000 System coupled to a nanoESI Q-Exactive HF mass spectrometer (Thermo Fisher Scientific).

The murine gut microbiota was analysed with a Q-Exactive HF mass spectrometer operated in nanoESI (1.6 kV). Full MS survey scans were recorded over the m/z range of 375-1500 with a resolution of 60 000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 60 ms. Up to 20 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 28%, with precursor isolation window at 2 m/z, resolution of 15 000, AGC value of 1E5 with a maximum injection time of 60 ms. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 sec.

4.2.4. Data analysis

Murine gut microbiome-specific protein databases were developed and used. These databases were not readily available as seen for model organisms (*E. coli*, *B. subtilis*, and *S. cerevisiae*) or standard mixtures (gut standard) seen earlier. Hence, for this study two approaches were taken.

Approach 1: Building database core group of bacteria at genus-level

Murine gut microbiome is known to be characterized by 37 genus-level core bacteria by using metagenomics approaches (Wang et al., 2019). This group of microorganisms was found in the faeces of 101 healthy murine differing in age, sex, and strain. In another study, the murine gut microbiome catalogue was developed by metagenomic sequencing of the faeces collected from 184 murine strains housed in various laboratories around the world. This comprehensive study provided metagenomes of the microbiome and also identifies a core set of 26 species (Xiao et al., 2015). So, for the first approach, a core set of 50 genus-level bacterial species along with host murine

were selected and their protein FASTA sequences were downloaded from the UniProt database. This list comprised the fifty most prominent bacterial species and their closely related genus protein sequence available in UniProt. The list of species is available in the annex. This database included 20 606 901 sequences with 8 689 718 023 residues.

Approach 2: Building a database from the MAGs provided by metagenomics

The study (Xiao et al., 2015) allowed for the development of MAGs (assembled genomes of the bacteria) from the gut microbiome isolated from 183 murine strains. The protein-coding regions and protein sequences can be predicted *in-silico* from these DNA sequences. Thus, the predicted protein sequence for the murine gut microbiome catalogue study was downloaded from MG-RAST accession ID 4661127.3 (<https://www.mg-rast.org/mgmain.html?mgpage=download&metagenome=mgm4661127.3>).

The downloaded sequences were translated and the protein sequence database was prepared with 2 189 553 sequences and 549 592 470 residues. The sequences are 10 times lower in quantity than approach 1.

Upon development of the two sequence databases, protein search was performed using proteome discoverer 2.4 with Sequest HT search engine with the conditions mentioned in section 5.2.2. Three searches with different processing workflows in PD with approach 1 database, approach 2 database, and combined 1 and 2 databases were used. Post-processing for functional and gene ontology was performed with UniProt KB and Perseus 2.0.3.0 software package.

4.3. Choice of the database

For peptide identification by using the Sequest HT search engine, it is necessary to understand the difference between them in terms of identification, proteome coverage, and accuracy. The comparison of the results obtained by combining all four replicates in terms of total identified proteins, peptides, PSMs, number of proteins with two or more peptides, number of proteins with one or more unique peptides, and number of proteins with one or more unique plus razor peptides (obtained from protein unique peptide in PD 2.4) is shown in table 5.

Table 5: Comparison of overall identifications obtained by different approaches used for murine protein database construction

	Total PSMs identified	Total peptides identified	Number of proteins identified with ≥ 2 peptides	Number of proteins identified with ≥ 1 unique peptide	Number of proteins identified with ≥ 1 razor + unique peptide	Total proteins identified
Approach 1	7134	4703	1278 (50 %)	2498 (98 %)	906 (36 %)	2543
Approach 2	7909	5337	1363 (52 %)	2615 (99 %)	1778 (68 %)	2631
Approach 1 + 2	9219	6241	1712 (51 %)	3236 (97 %)	1429 (43 %)	3330

As expected, the numbers of total identified proteins (3330 proteins), total peptides (6241 peptides), and total PSMs (9219) that indicate the proteome coverage are the highest when a combination of approaches 1 and 2 was used for database building. However, it is interesting to note that despite having a 10 times smaller sequence-sized database, approach 2 has better identification than its counterpart approach 1. The impact of the database size has been explained before to be advantageous to choose one with a small size and with the actual sequences expected to be identified to avoid false PSMs and reduced FDR values (Jeong et al., 2012).

Further, the level of trust in protein identification between the methods can be understood by the peptide identification corresponding to each protein. The conditions like protein identification with more than two peptides or proteins with at least one unique peptide will determine the level of trust of the protein match after peptide identification, but this condition does not have enough accuracy in predicting the protein's species of origin. Whereas, the condition to have one *razor peptide* (peptides that are found in proteins that have the same homology) and a unique peptide (a peptide, irrespective of its length, that exists only in one protein of a proteome of interest it can however be identified more than once in the entire sequence) further proves the association of the protein to a specific species proteome. This is a major issue in metaproteomics, as the microorganisms can be genetically very closely related to each other. The number of proteins with more than two peptides or at least one unique peptide is shown to be almost similar between the three methods. However, when the estimation of the combination of razor peptides and unique peptides was made, approach

2 had twice the identification than other methods used. This shows that approach 2 containing protein sequences developed from MAGs has a higher chance of identifying the species responsible for producing the protein. This is an additional reason to perform metagenomic sequencing on the sample before, or at the same time, metaproteomic analysis.

The results validate that higher accuracy in protein identification is achieved when the search is done with a protein sequence database built using approach 2. Also, to facilitate the building of the database described by approach 2 and also perform proteomic studies, it will be necessary to perform multi-omics and again the sample preparation may be made easier with ChipFilter workflows. I wish to remark that sample preparation strategies will also play an important role in the data analysis results and the selection of a robust technology like the ChipFilter method can be highly beneficial.

4.4. Taxonomic analysis

Although it was explained that using a combined approach or only approach 2 to be beneficial, I had to use approach 1 for the subsequent analysis as the protein annotation for approach 2 was not fully achieved before the completion of the thesis and is still in progress. Protein annotation is a major issue for database quality and should be taken into account in the choice of database.

ChipFilter allowed the identification of 810 ± 122 proteins, 2310 ± 296 peptides, including 1372 ± 165 unique peptides, and 4394 ± 506 PSMs (table 6). It is interesting to note that almost 98% of the proteins identified contain at least one unique peptide, indicating high confidence in identification.

The ChipFilter workflow has shown to be well suited in this case to prepare the highly complex murine faecal microbiome to identify species from different phyla. Baniasad M. *et al.* (Baniasad et al., 2022) compared different lysis buffers and preparation strategies with FASP, SP3, and S-Trap for murine faecal metaproteomic sample preparation to identify proteins between a range of 1600 to 3000 for different lysis buffers for S-Trap method and 750 to 1800 proteins for SP3 method. In comparison to this, the protein identification by ChipFilter workflow is respectable as well, especially considering that the analytical methods (LC gradients, mass spectrometer acquisition) have not been specifically optimized for these analyses.

Table 6: Number of identified proteins, peptides, PSMs, and unique peptides of four major phyla for four samples processed by ChipFilter workflow

	Number of Proteins	Number of Peptides	Number of PSMs	Number of Unique Peptides
Firmicutes	536 ± 136	1490 ± 261	2742 ± 411	824 ± 157
Bacteroidetes	111 ± 50	309 ± 144	550 ± 236	167 ± 72
Proteobacteria	78 ± 47	225 ± 137	352 ± 206	137 ± 81
Actinobacteria	13 ± 2	37 ± 9	106 ± 31	19 ± 4
Others	74 ± 15	249 ± 32	645 ± 60	225 ± 30

Phylum

The results we obtained in the present analysis for murine gut microbiome identify the major phylum like Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Verrucomicrobia, Cuneatibacter, Cyanobacteria, and Mycoplasmota. Firmicutes are a phylum of bacteria, most of which have gram-positive cell wall structures (Hashmi et al., 2020). This phylum in the gut microbiota has been demonstrated to be involved in energy resorption, and potentially related to the development of diabetes and obesity (Ley et al., 2006). The phylum Bacteroidetes comprising of three large classes of Gram-negative bacteria that are widely found in the environment, including in soil, sediments, and seawater, as well as in the guts and on the skin of animals (Flint & Duncan, 2014). The phylum Proteobacteria consists of a wide variety of pathogenic genera. Some are free-living (non-parasitic) and include many of the bacteria responsible for nitrogen fixation (Rizzatti et al., 2017). Actinobacteria are one of the four major phyla of the gut microbiota representing only a small percentage and are pivotal to maintaining gut homeostasis (Binda et al., 2018).

In figure 43, the phylum Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria accounted for 62%, 12%, 9%, and 2% on an average between four samples respectively. While the remaining 15% comprises, other phyla including murine proteins. The report suggests that the phylum of bacteria identified by ChipFilter workflow is also commonly identified in the murine gut (shown by genomics) but with different compositions depending on the physical state of the murine (Chu & Zhang, 2022). The results revealed that at the phylum level, dominant bacteria in the gut microbiota of murine mainly included Firmicutes (55.75%), Bacteroidetes (37.02%), Proteobacteria (4.05%), Actinobacteria (1.98%), and Tenericutes (1.09%), while the abundance of other bacteria was less than 1%. My results are also in line with the results represented in figure

24 for species identified per phylum as seen in humans. All phylum except fusobacteria was identified in the murine gut microbiome.

It can also be seen that the phylum abundance is highly dissimilar between the samples. While sample 1 is enriched with Firmicutes, sample 2 seemed to have more Proteobacteria than other samples, sample 3 is enriched by Bacteroidetes as compared to others and sample 4 contains relatively higher Actinobacteria and other phyla like Verrucomicrobia, Cuneatibacter, Cyanobacteria or Mycoplasmota.

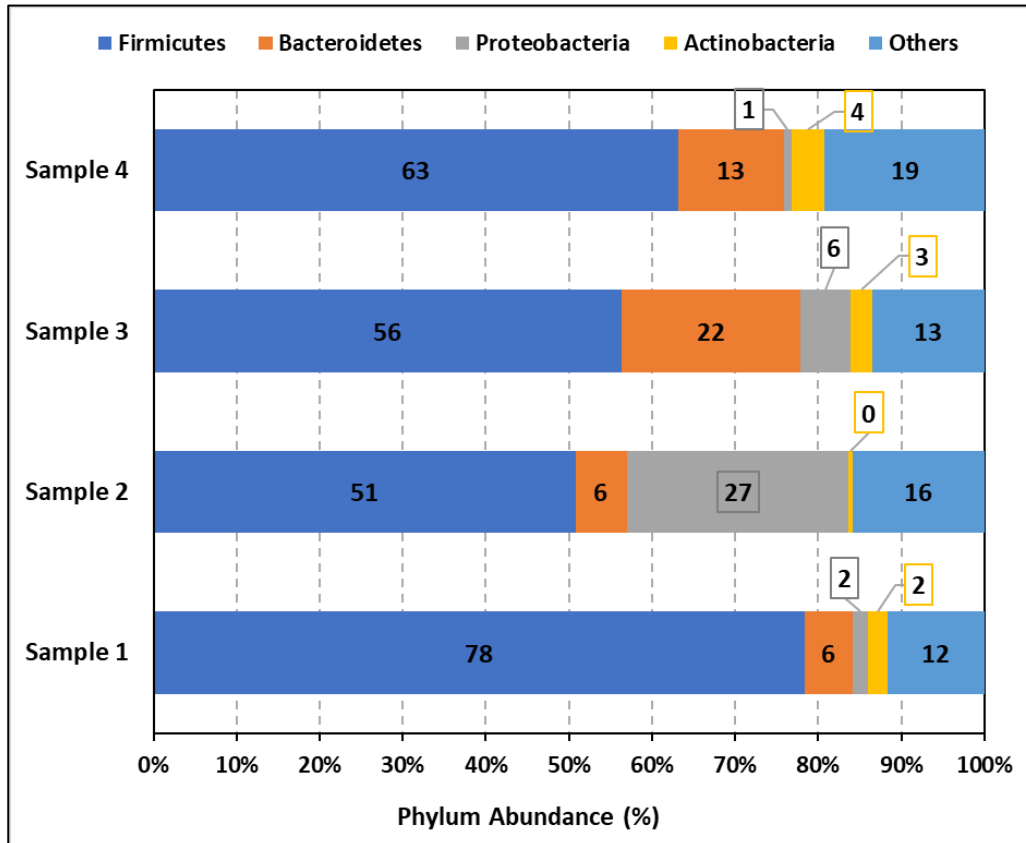


Figure 43: Phylum abundance in different murine faeces samples processed by ChipFilter workflow estimated by Kleiner et al., 2017

Genus

A core set of 37 genera of microorganisms has been identified from the four processed samples by ChipFilter along with murine proteins identified in all the samples. The list of genera includes- *Achromobacter*, *Alistipes*, *Anaerocolumna*, *Bacillus*, *Bacteroides*, *Blautia*, *Butyrivibrio*, *Clostridium*, *Coprococcus*, *Dorea*, *Epulopiscium*, *Erysipelotrichaceae*, *Escherichia*, *Eubacterium*, *Extibacter*, *Intestinibaculum*, *Lachnobacterium*, *Lachnoclostridium*, *Lachnospiraceae*, *Lacrimispora*, *Lactobacillus*,

Lactococcus, *Limosilactobacillus*, *Paeniclostridium*, *Parabacteroides*, *Prevotella*, *Pseudobutyrvibrio*, *Pseudomonas*, *Rhizobium*, *Roseburia*, *Ruminococcaceae*, *Ruminococcus*, *Schaedlerella*, *Subdoligranulum*, *Suipraeopectans*, *Tannerella*, and *mus*. This includes twenty-seven Firmicutes, five Bacteroidetes, four Proteobacteria, and one murine. The distribution of different species between the four samples is shown in figure 44 along with the 37 genera identified from all the sample replicates. Twenty of the genera identified were found to be reported as the core set of bacteria in the normal murine by genomic sequencing (Chu & Zhang, 2022).

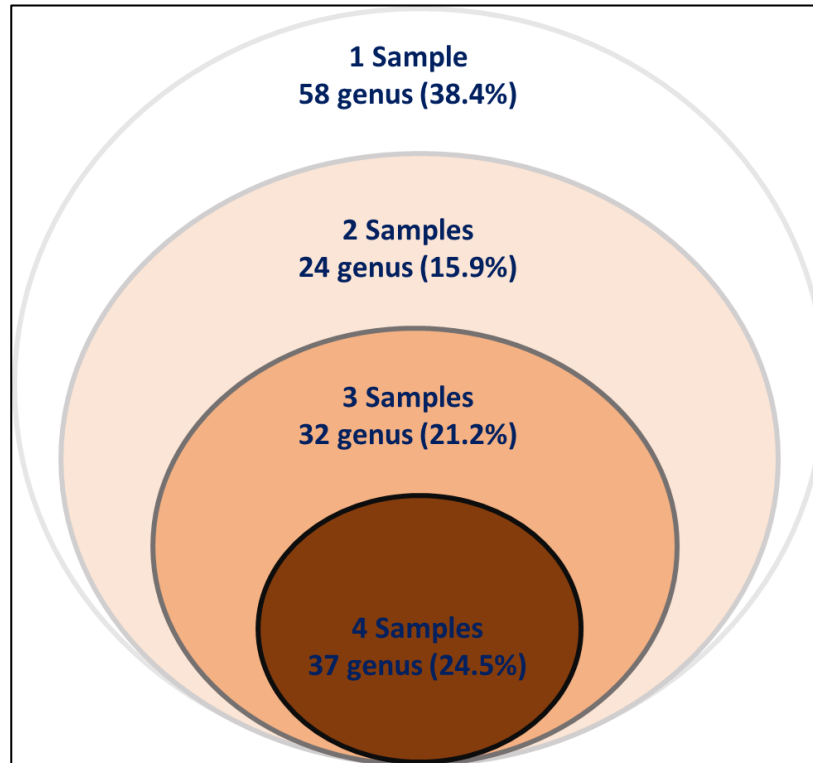


Figure 44: Genus's level distribution between different murine faeces samples processed by ChipFilter workflow

4.5. Functional analysis

The proteins identified from the four murine samples after processing by ChipFilter workflow were functionally annotated. Out of the total 2543 proteins identified, 1576 (62%) proteins were assigned for biological process, 1304 (51%) were assigned for cellular components and 2228 (88%) were assigned for molecular function. The results are shown below.

Biological process

A total of 473 biological processes were associated with the identified proteins. Of them, the top 20 biological processes with the respective counts are shown in figure 45. Ten out of the twenty processes are metabolic processes, four in central dogma, and three each in transport and protein modifications. The identification of a greater proportion of metabolic proteins means that the contribution of the microbiome to host food digestion is significant, thereby maintaining a healthy symbiosis. This indicates that the microbiome is active in metabolism and replication inside the host. Interestingly, some pathways associated with antibiotic resistance have been also identified and they correspond to proteobacteria like *Escherichia*, *Pseudomonas*, and *Achromobacter*.

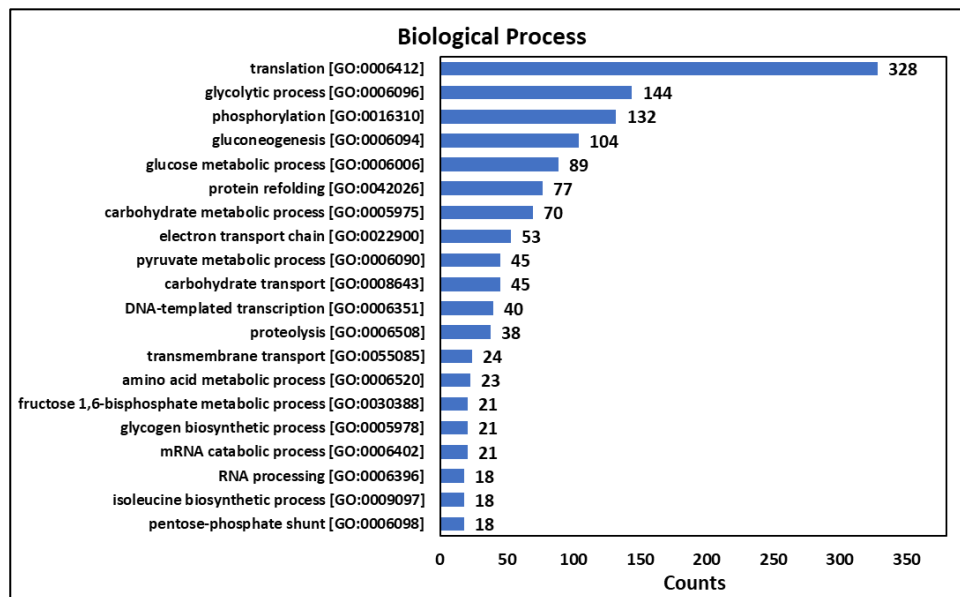


Figure 45: Top 20 biological processes seen for the identified proteins from murine gut microbiome

Cellular compartment

A total of 146 cellular components were associated with the identified proteins. The top 20 cellular components include cytoplasm, extracellular, membrane, cell surface, and ribosomal complex (figure 46). This shows that the distribution of identified proteins is uniform to different sites of the bacterial and murine cells. This is essential to validate the sample preparation process, particularly the cell lysis and protein solubilisation steps.

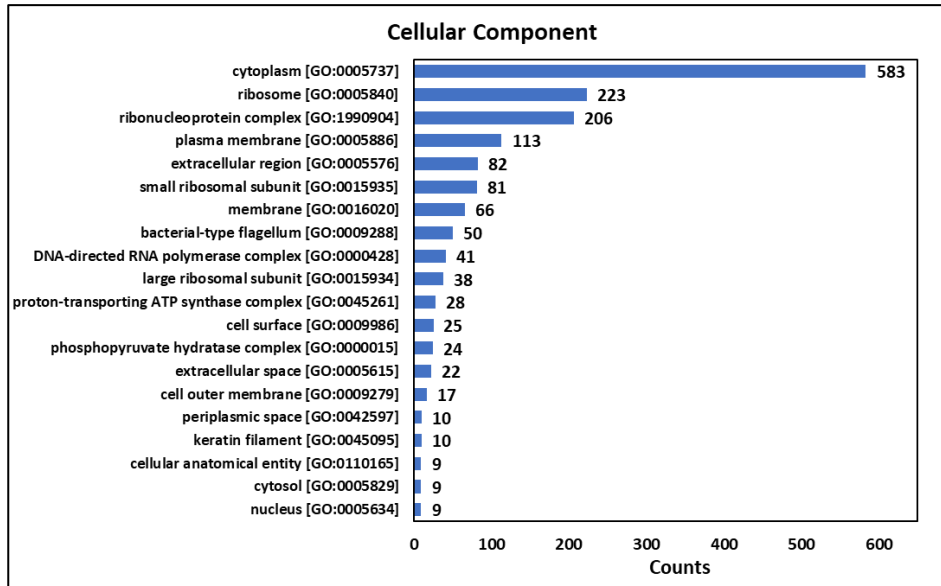


Figure 46: Top 20 cellular localisations seen for the identified proteins from murine gut microbiome

Molecular function

A total of 500 molecular functions were associated with the identified proteins. The top 20 molecular functions include enzymatic activity and binding activity of energy molecules like ATP, GTP, and NADP (figure 47).

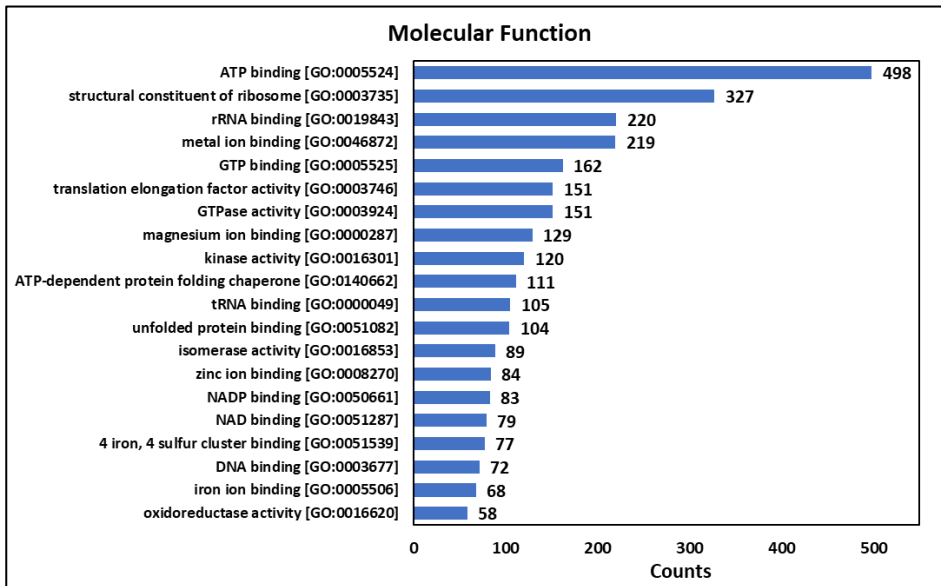


Figure 47: Top 20 molecular functions seen for the identified proteins from murine gut microbiome

4.6. Conclusions

An unknown mixture of microorganisms obtained from murine faeces corresponding to the gut community was processed by the ChipFilter method for bottom-up proteomics.

To identify the peptide sequences from MS acquisition, a protein sequence database was constructed by using a core group of bacteria at the genus level and from the MAGs provided by metagenomics. A comparison of initial identification obtained by using the two databases showed that the metagenomic developed database provided better reliability in matching the protein identified to the species of origin in a sample community. This result proves that metaproteomics is better resolved when it is accompanied by metagenomic sequencing.

This showed that multi-omics studies are necessary and having ChipFilter-based sample preparation for multi-omics can be very beneficial.

The phylum Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria accounted for a majority of the species identified with different phylum abundance between samples.

A core set of 37 genera was identified as common to all the samples processed.

5. Going beyond proteomics

5.1. Introduction

Until now I have established the application of the ChipFilter in sample preparation for metaproteomics that includes all the steps from cell lysis to proteolysis. Further, to expand the usage of the ChipFilter, I will describe in this chapter the usage of ChipFilter to isolate nucleic acids from the same sample that was used for proteomics sample preparation (figure 48). This novel approach to isolating protein and nucleic acid contents from the same sample allows for proteogenomics studies very accurately.

As a first step, workflow 4 for the isolation of the nucleic acid isolation after the proteolysis is explained in detail. Next, the validation of the isolated nucleic acid is done by agarose gel electrophoresis for microbial and human cell mixtures. Finally, the amplification of the isolated nucleic acids (plasmid DNA) by polymerase chain reaction (PCR) was done.

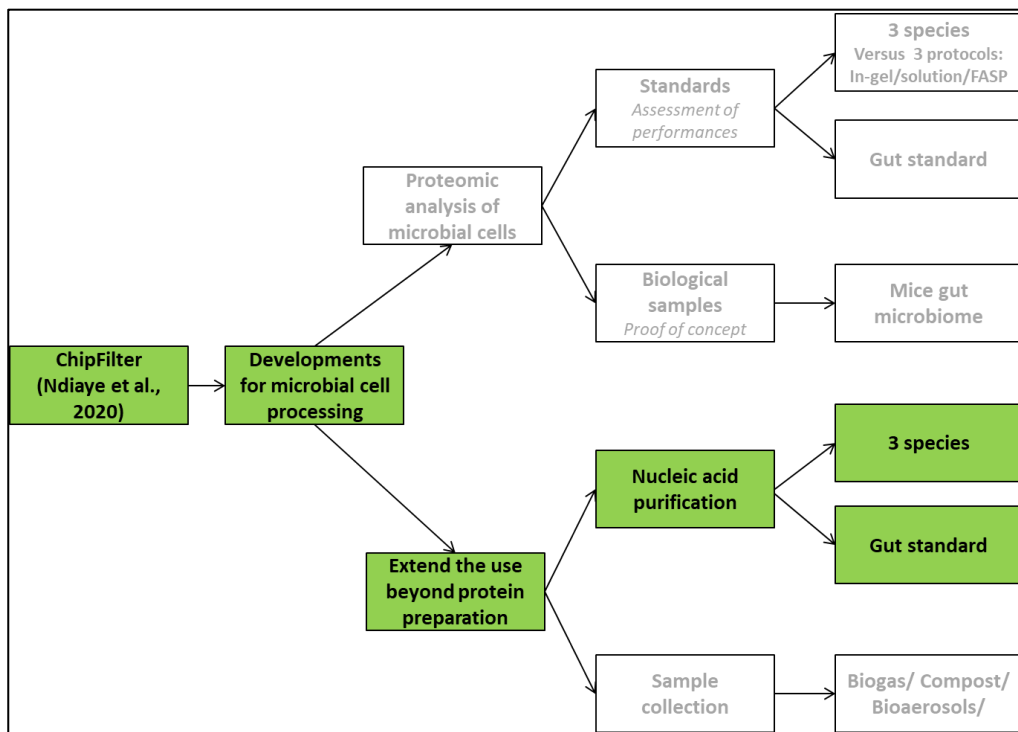


Figure 48: Steps discussed in chapter 5

Sample preparation workflows demand more than single biomolecule isolation as the major bottleneck in multi-omics is the limitations in available sample preparation workflows. Additionally, as seen before, studies on microbiomes require the integration of more than one omics study. In that regard, an effort is made to understand the sequential isolation of nucleic acids after protein digestion in the ChipFilter workflow.

The inspiration for this part of the study was from earlier work published (Wiśniewski & Rakus, 2014) to show the application of multi-enzyme digestion FASP for absolute quantification of *E. coli* proteome. The authors establish multi-enzyme digestion of proteins, RNA, and DNA using proteases (Lys-C and trypsin) and nucleases (RiboShredder and DNase) to sequentially remove peptides, RNA, and DNA fragments. This demonstration in FASP allowed me to experiment with the ChipFilter workflow to isolate sequentially the different biomolecules. Unlike FASP, the ChipFilter offers the advantage of bidirectional flow i.e., the samples can be eluted from both ways to the reaction chamber by alternating the inlet flow direction of the solution. This allows the simple removal of the nucleic acids left behind after proteolysis.

5.2. Material and methods

5.2.1. Biological material

The cells used in the study include gut standard microbe mixture (a mixture of 21 strains of gut microorganisms in variable composition obtained commercially and is discussed in chapter 5) (Zymo Research), human adenocarcinoma cancer cells HT29-MD2, individual and mixed cultures of the microbes - *E. coli* expressing plasmid pGEX, *B. subtilis*, *S. cerevisiae*.

5.2.2. Nucleic acid purification

As described in section 2.4 the nucleic acids were collected after proteolysis in ChipFilter using (workflow 4). The parameters specifically for nucleic acid recovery are discussed in the following section.

For comparison with a standard method, TRIzol (Thermo scientific) protocol for RNA/DNA isolation was used, QIAprep spin miniprep kit (Qiagen) workflow was used for *E. coli* plasmid isolation, and ZymoBIOMICS DNA Microprep Kit for gut standard DNA isolation. All these protocols were realized according to the manufacturer's recommendations. For protocols, please refer to the following links

- TRIzol protocol

https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf

- QIAprep spin miniprep kit protocol

https://molbio.mgh.harvard.edu/szostakweb/protocols/clone_seq/qpmini_spin.pdf

- ZymoBIOMICS DNA Microprep kit protocol

https://files.zymoresearch.com/protocols/d4301_d4305_zymbiomics_dna_microprep_kit.pdf

5.2.3. Validation by agarose gel electrophoresis

Nucleic acids were analysed on 1% or 1.5% agarose gel stained with GelRed.

5.2.4. Validation by Polymerase Chain reaction

PCR amplification was done with the following conditions in a thermocycler (Analytik Jena) with an initial temperature of 95 °C for 2 minutes and 30 cycles with 92 °C for 30 seconds, 55 °C for 15 seconds, 72 °C for 30 seconds. After a brief heating at 72 °C for 10 minutes was done and products were stored at 4 °C until gel electrophoresis. The experiment was performed in 3 replicates and positive control of the purified plasmid was taken. The initial cell density of *E. coli* used was $\sim 3 \times 10^7$.

PCR was done on DNA extracted in the device from $\sim 3 \times 10^7$ *E. coli* cells transformed with a GFP containing pGex 4-T3 plasmid. The PCR was done using 2.5 units Taq polymerase (Qiagen) with forward primer (5' TCC CCG AAT TCT ATG AGT AAA GGA 3') and reverse primer (5' AGG GGG AAT TCT TAT TTG TAG AGC 3') synthesized by Eurogentech. Thermal cycling was done with the following conditions in a thermocycler (Analytik Jena) with an initial temperature of 95 °C for 2 minutes and 30 cycles with 92 °C for 30 seconds, 55 °C for 15 seconds, 72 °C for 30 seconds. After a brief heating at 72 °C for 10 minutes was done and products were stored at 4 °C. PCR products were separated by electrophoresis on 1% agarose gel stained with GelRed.

5.3. When do the nucleic acids leave the ChipFilter?

It is imperative to know the different stages at which the nucleic acids are removed from the ChipFilter system. Cell numbers utilized for proteomic studies by ChipFilter workflow can range from a few hundred to a few million, but the nucleic acid amount obtained is not entirely proportional to the initial cell density as the loss of nucleic acid during different steps of the proteomic workflow.

To assess the loss of nucleic acids during sample preparation, a gut standard mixture (approximately 4×10^8 cells according to the manufacturer) has been loaded on the ChipFilter device and submitted to the steps before proteolysis. The flow through was collected after each step, from cell loading to proteolysis, has been recovered. The nucleic acids were recovered by inverting the flow (WF4). All the samples were analysed on agarose gel (figure 49).

Results indicate that a greater fraction of the nucleic acid component was removed during cell loading, the abnormal migration pattern of the nucleic acids indicates that it has interference from cellular components that alter the migration of the nucleic acids within the gel. Next, a small portion of the nucleic acid was also washed during the cell lysis buffer step. The reduction, alkylation, and washing steps do not have any nucleic

acid loss. Finally, during the elution step 1 (55 ng/μl for DNA estimation by nanodrop), better quality and quantity of nucleic acid components as seen by intense bands were obtained. But, the quantity estimation by nanodrop and also seen in the gel is lower than cell loading (777.5 ng/μl for DNA estimation by nanodrop) but the quality is much better. The value of 260/280 was 1.19 for cell loading, while it was 1.55 for elution step 1. The same experiment was also performed with mixed cells (mix 1) and similar results were obtained (data not shown).

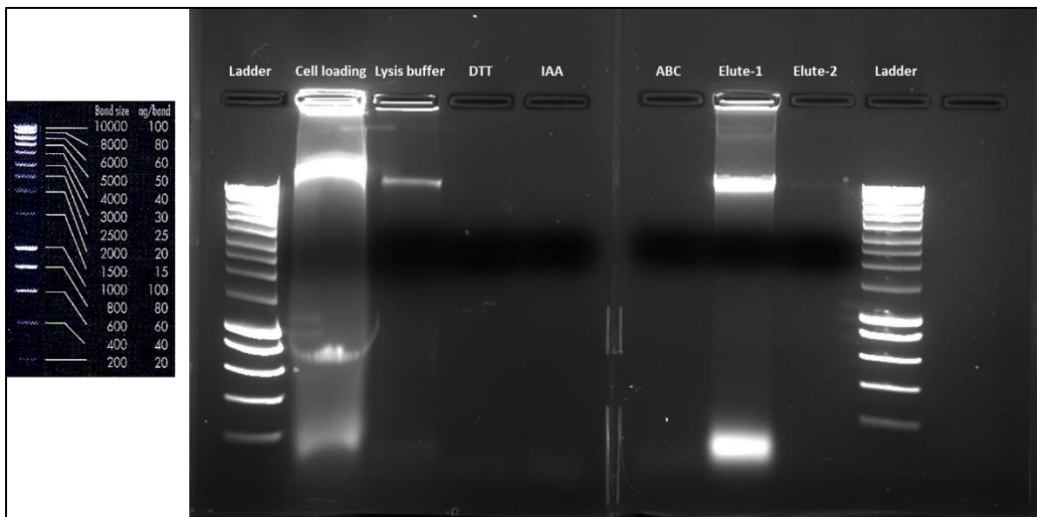


Figure 49: Monitoring the removal of the nucleic acids at different steps of the ChipFilter WF4

Lanes from left to right contain a ladder, flow through collected during cell loading, lysis buffer loading, reduction (DTT), alkylation (IAA), (ABC), first elution by reversing the flow, second elution, and ladder. This experiment was performed with a gut standard microbe mixture which is approximately 4×10^8 cells. 35ul of each of the fractions was resolved in 1% (m/v) agarose gel electrophoresis with a voltage of 120V.

5.4. What is the minimum number of cells required to obtain nucleic acid by ChipFilter workflow?

It is necessary to know the amount of starting material in terms of the number of cells to estimate the nucleic acids and proteins that can be simultaneously useful for nucleic acid analysis by PCR or NGS and MS respectively. As seen, a few hundred microbial cells were sufficient to identify more than 500 proteins but this may not be the case for nucleic acid due to the loss seen before. Also, the subsequent analysis or nucleic acid identification approach will determine the requirement.

Human cells (10^6) and microbial cells (mix 1: 10^6 of *E. coli*, *B. subtilis*, and *S. cerevisiae*) were loaded on the device. After proteomic sample preparation, nucleic acid

was recovered as previously described and analysed on agarose gel (figure 50). In both cases, nucleic acids could be identified. When compared to the commercially available methods for DNA extraction from bacteria, it was shown that the recommended number of cells was between 10^7 to 10^9 cells (Becker et al., 2016).

Additionally, when the gut microbiome standard (10^8) was used to extract nucleic acids as shown in figure 44, the intensity yield was higher than for 10^6 human or mixed cells. When compared to another study reporting the DNA yields from different methods dedicated to processing faeces, it was reported to have the highest yield of 12 μg (Gryp et al., 2020). Using the gut standard composed of the same microbes identified in the human gut (explained in chapter 6) in ChipFilter, the yield identified by nanodrop estimation was approximately 2 μg . This yield is compatible with further analysis by PCR or NGS.

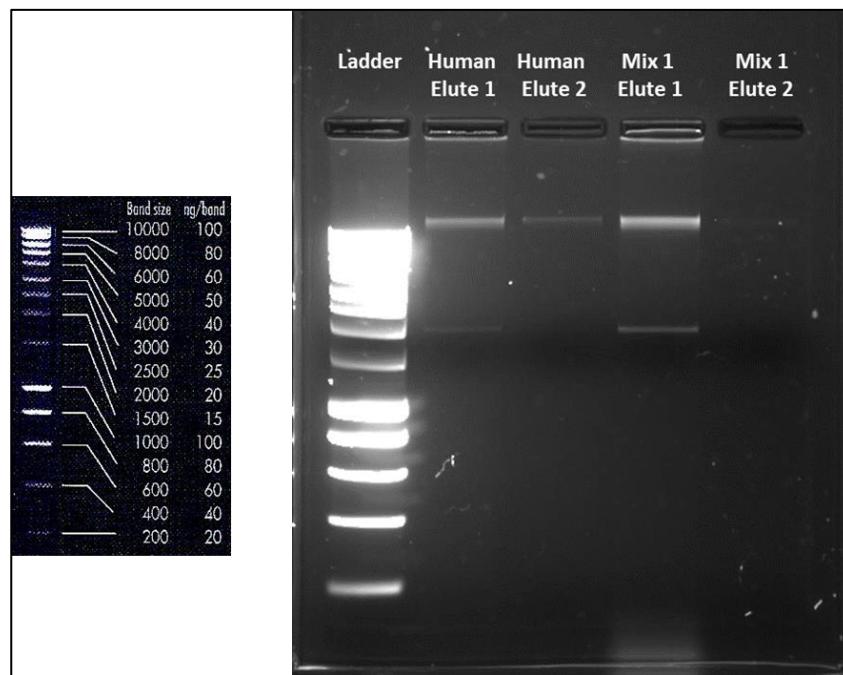


Figure 50: Minimum human (10^6) and microbial (mix 1) cell number necessary for visualization in agarose gel electrophoresis using ChipFilter WF4

Lanes from left to right contain a ladder, first and second elution by reversing the flow for human cells (1×10^6 cells) and mix 1 (3×10^6 cells). 35 μl of each of the fractions was resolved in 1% (m/v) agarose gel electrophoresis with a voltage of 120V.

5.5. How are the nucleic acids retained by the ChipFilter?

It is noticed that nucleic acid exit from the ChipFilter during cell loading and lysis step. Also, the proportion of injected cell number does not contribute exactly to the

obtained nucleic acid obtained and there is a loss. Hence, it is necessary to understand the mechanism of the nucleic acid trapping inside the ChipFilter.

To understand the process of nucleic acid removal, DNA and RNA fractions isolated from 10^6 human cells were injected into the ChipFilter and the flowthrough was collected. A subsequent washing step was performed once and the flowthrough was collected. The collected flowthrough of 30 μ l was completely resolved by agarose gel electrophoresis and the results are shown in figure 51.

It was observed that the nucleic acids did not bind or get trapped inside the ChipFilter causing them to be eluted at the beginning step while they are injected. This explains why the DNA is eluted during loading and lysis steps in the earlier shown cases.

A possible explanation for this phenomenon is that only the protein-bound nucleic acid materials are retained while all the unbound materials are removed during injection or within the very first or second step of rinsing. Further, this theory is supported by the fact that the proteomic analysis of all the samples subjected to workflow-4 identified DNA binding proteins of prokaryotic and eukaryotic origin. The prokaryotic-specific DNA-binding proteins that were identified include DNA-binding protein HU, Integration host factor, ribosomal proteins, protein complexes in translation, DNA replication, and DNA protection. The eukaryotic-specific DNA binding proteins that were identified include histones, DNA replication, transcription, and RNA binding proteins forming translation complexes.

The implication of this theory would mean that there is a selective enrichment of the nucleic acid regions that are bound to a known set of proteins. On the contrary, nucleic acids fraction that is unbound will be lost in the process and hence a thorough study would require isolating the fraction removed during injection and washing/cell lysis. I am interested to understand the nucleic acid sequences that will be obtained as a result of performing NGS on the eluted fraction using ChipFilter workflow-4. The samples for sequencing have already been given, but obtaining the sequencing results and data analysis will take more time and so this hypothesis is yet to be completely proved. The focus of this section is to establish the idea that nucleic acids are automatically released without inhibition inside the ChipFilter. In the case where they are isolated as a cellular fraction, retention is observed. This phenomenon can be described as a result of the retention of DNA- or RNA- binding proteins.

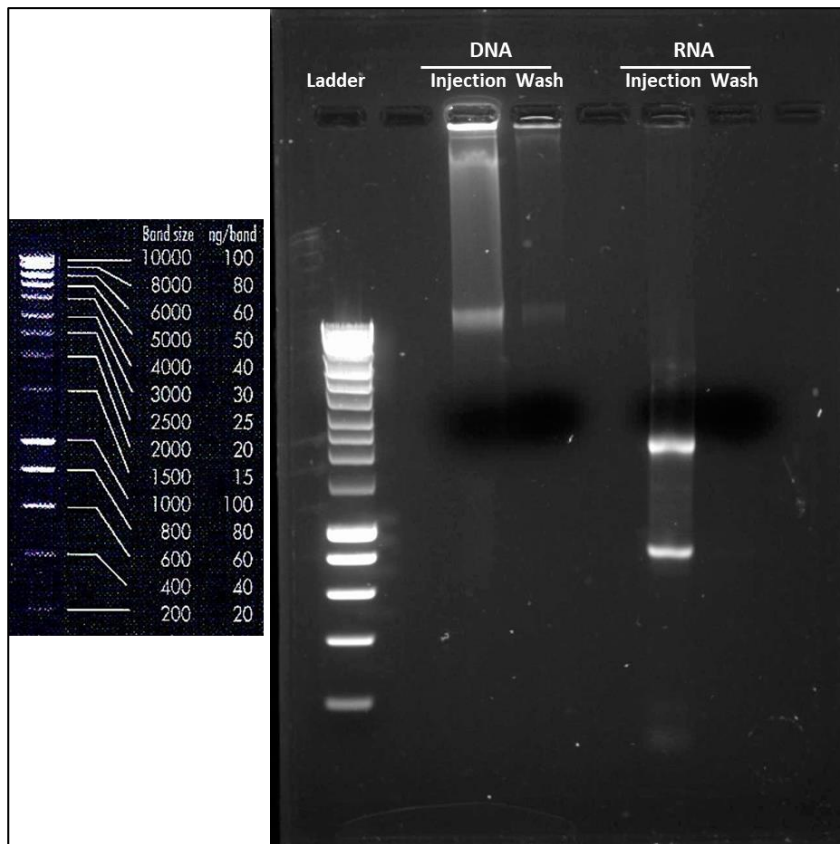


Figure 51: Representation of the DNA and RNA retention onto ChipFilter

Lanes from left to right contain a ladder, flowthrough collected during injection and washing steps for DNA and RNA from human cells (1×10^6). 30 μ l of each of the fractions was resolved in 1% (m/v) agarose gel electrophoresis with a voltage of 120V.

5.6. Do the reagents used in proteomic sample preparation affect the nucleic acids?

The elution of the nucleic acid is positioned to be after proteolysis which will involve a plethora of reactions with chemicals like non-ionic detergents, DTT, Iodoacetamide, and enzymes that may also affect the nucleic acid present in the system. Hence, it is necessary to ensure they have been well protected from these chemicals.

As a first step, the effect of these chemicals was determined by incubating the DNA and RNA isolated by the TRizol method from 1×10^6 human cells with 20ul of each lysis buffer I, 20 mM DTT in lysis buffer I, 50 mM IAA in lysis buffer I, 50 mM ABC buffer and 2 μ g trypsin in a vial at room temperature for 1 hour in parallel. Positive control of the DNA or RNA was kept at room temperature for the same time as the test condition was taken.

The results for DNA (figure 52) and RNA (figure 53) indicate that there is no detectable effect as the migration seems not affected upon treatment with the respective reagent. This must be further confirmed by sequencing experiments or by PCR. No proof showing the deleterious effect of non-ionic detergent or IAA was found, but a study on the effect of DTT on double-stranded plasmid showed above 10 mM, nicks on the plasmid are introduced as a result of exposure (Fjelstrup et al., 2017). This could not be confirmed in my study and hence further study might be necessary. As for now, no effect on the nucleic acids by the proteomics reagents was observed.

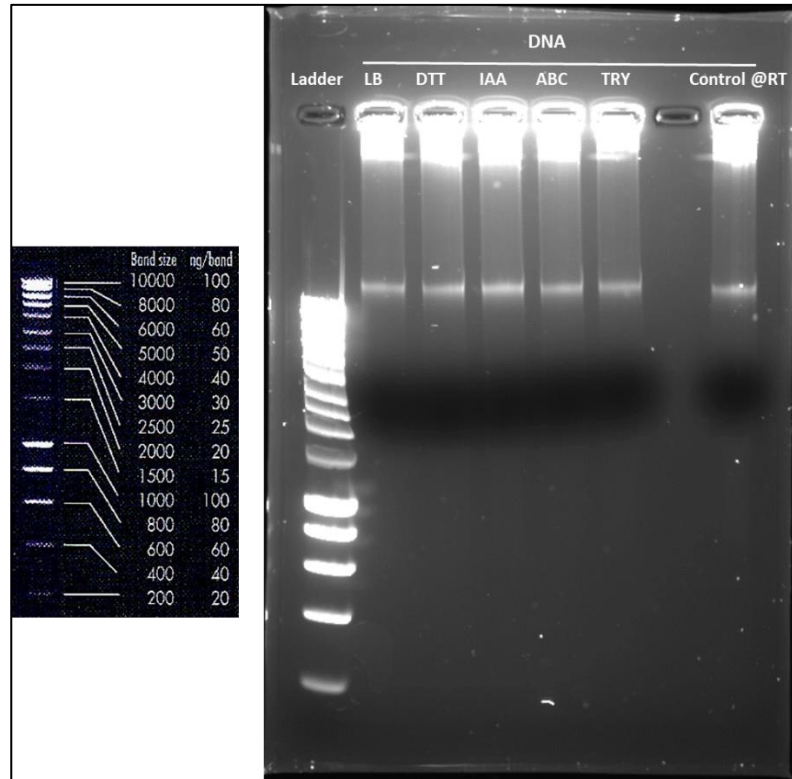


Figure 52: The effect of proteomic reagents on DNA

Lanes from left to right contain a ladder, DNA incubated with lysis buffer (LB), DTT, IAA, ABC, Trypsin (TRY), and control without treatment. 10 μ l of each of the fractions was resolved in 1 % (m/v) agarose gel electrophoresis with a voltage of 120V.

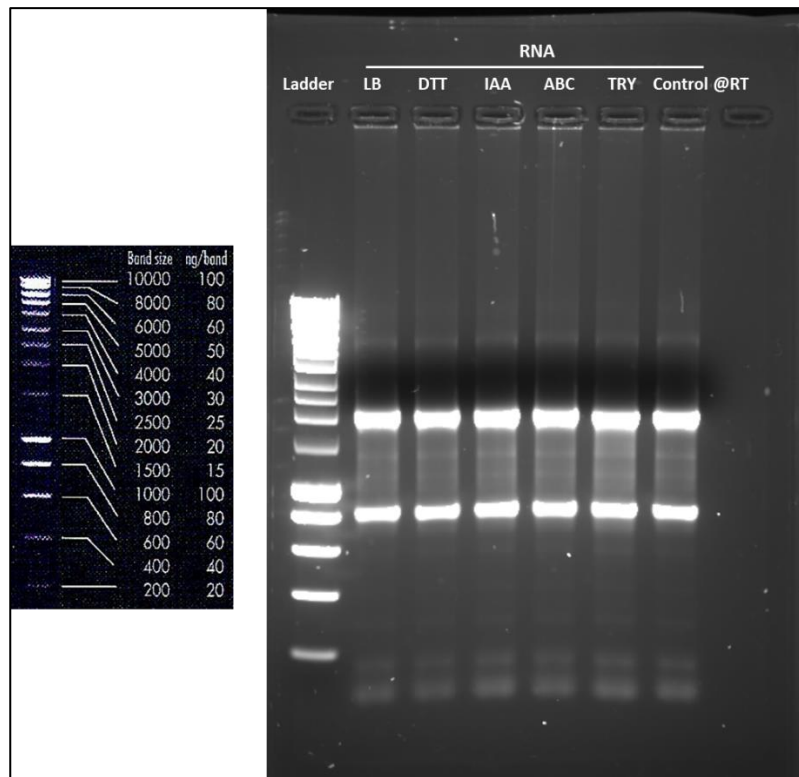


Figure 53: The effect of proteomic reagents on RNA

Lanes from left to right contain a ladder, RNA incubated with lysis buffer (LB), DTT, IAA, ABC, Trypsin (TRY), and control without treatment. 20 μ l of each of the fractions was resolved in 1 % (m/v) agarose gel electrophoresis with a voltage of 120V.

5.7. Is the DNA recovered compatible with PCR analysis?

To functionally validate the integrity of the nucleic acids isolated by ChipFilter WF4 and to demonstrate the possibility of downstream analysis by PCR, I chose first to amplify a specific region of a plasmid that was present in the *E. coli* strain used in the study. For this purpose, the green fluorescent protein-encoding region of the pGEX-4T-3 plasmid was targeted, leading to a fragment of 780 bp.

For all three replicates, a single product at 780 bp is formed (figure 54). The intensity of the band is diminished as compared to PCR prepared on the pure plasmid but can be identified. This confirms that the nucleic acid fraction contains plasmid DNA and it can be functionally validated using the PCR experiments after isolation using ChipFilter. This validation is critical to lead to the possibility of multi-omics sample preparation using ChipFilter.

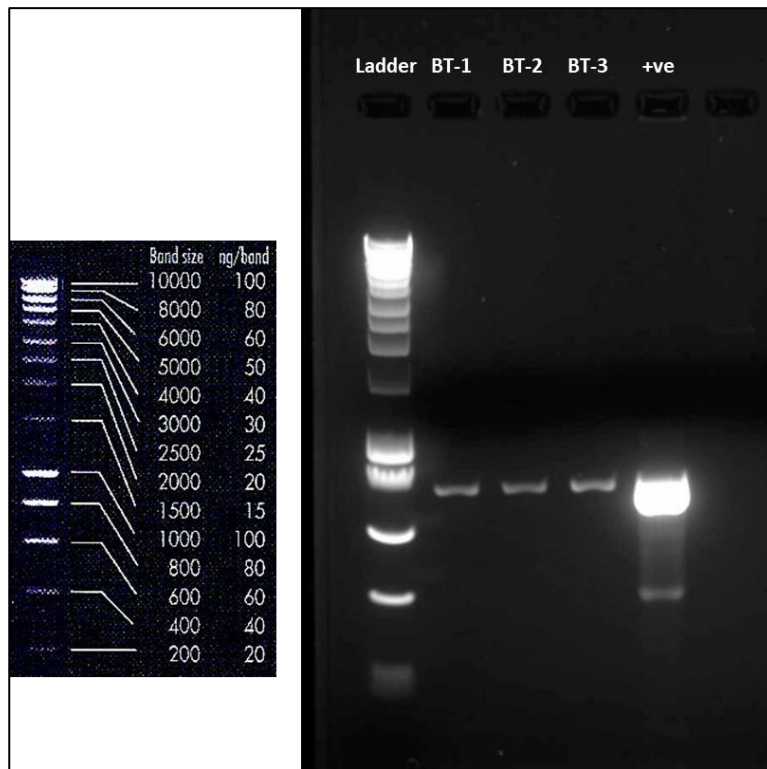


Figure 54: PCR products formed from the amplification of plasmid DNA isolated from *E. coli* with ChipFilter WF4

Lanes from left to right contain a ladder, PCR products for three replicates (BT-1, BT-2, and BT-3), and positive control (+ve). 10 μ l of each of the fractions was resolved in 1.5 % (m/v) agarose gel electrophoresis with a voltage of 120V.

5.8. Conclusions

The study of nucleic acids on the same samples used for proteomic analysis would be a very promising advance for multi-omics analysis, allowing the reduction of sampling bias. To achieve this objective, various steps have been taken:

- ✓ Nucleic acids can sequentially be removed from cells used for proteomics sample preparation with WF4 in ChipFilter.
- ✓ The nucleic acids are found to exit the ChipFilter during cell loading, cell lysis, and elution with the reversed flow direction.
- ✓ To be visualized by agarose gel electrophoresis, a minimum of 10^6 cells either of prokaryotic or eukaryotic origin has to be taken. This is necessary to overcome the loss that happens during cell loading and lysis.
- ✓ The holding of the nucleic acids within the ChipFilter is hypothesized to be indirectly mediated by the retention of nucleic acid-bound proteins such as DNA-binding proteins, RNA polymerase, and histones.

- ✓ The nucleic acids (DNA and RNA) were found to be stable after incubation with lysis buffer I, 20 mM DTT, 50 mM IAA, 50 mM ABC buffer, and 2 µg trypsin.
- ✓ The nucleic acid extraction was performed from cells of (i) human origin, (ii) a defined mixture of *E. coli*, *S. cerevisiae*, and *B. subtilis*, and (iii) a gut microbiome standard mixture comprising 21 different species of bacteria, fungi, and archaea.
- ✓ Functional validation of the isolated nucleic acid fraction was done by using PCR amplification of a specific region in the plasmid isolated from *E. coli* cells by ChipFilter WF4.

As a future study, PCR amplification dedicated to other nucleic acid fractions like DNA and RNA is planned with the primers specific for *E. coli* and *S. cerevisiae* that have been designed and obtained. Moreover, simultaneous NGS sequencing and proteomics sequencing of gut microbiome standard will be introduced in section 6.2.

6. Toward a multi-omics approach for studying a microbiome standard of the gut

6.1. Introduction

The demonstration of the ChipFilter workflows for proteomic and nucleic acid isolation has been done. A simple mixture of microbial cells with varying cell density was also used to show the difference in sample preparation between ChipFilter workflows and other commonly used workflows. Nucleic acid extraction after proteomics sample preparation was realized on a simple mixture of microorganisms.

However, it is well known that the microbiome will not have the same cell number throughout and that the difference in cell number between the participating species is a characteristic trait. Furthermore, the composition of any microbiome will include a variety of species belonging to bacteria, protozoa, viruses, archaea, or eukaryotes (fungi and algae). A good sample preparation workflow should enable the processing of most of the species to increase identification coverage and reduce identification bias. This ideal workflow must be capable of handling a low cell number population and as well lyse/recover proteins and nucleic acid from different species having structurally complex cell characteristics.

To show the robustness of this workflow, it will be used on a mixture of microbial cells that have been reported to be identified from the gut microbiome of humans. The use of a commercial standard has the advantages of increasing the diversity of the microorganisms studied, and of knowing the exact composition of the microbiome analysed, from both qualitative and quantitative points of view. This chapter will show the workflow for simultaneous sample preparation of gut standards for proteomics and genomics. The results are highlighted with the peptide-level identifications, species-level distribution, and nucleic acid quality and quantity evaluation (figure 55).

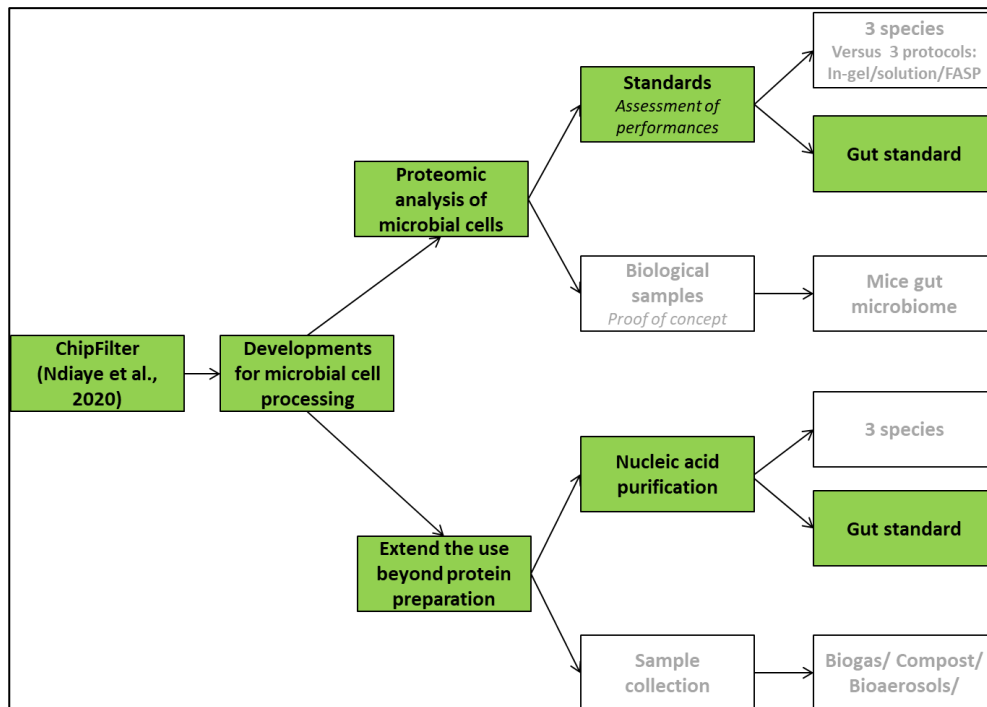


Figure 55: Steps discussed in chapter 6

6.2. Material and methods

6.2.1. Biological material

A standard whole-cell mixture consisting of 21 representative strains from 17 representative species of the gut microbiota (Zymo Research, ref D6331) was divided into 10 aliquots and stored in the storage solution provided by the manufacturer at $-80\text{ }^{\circ}\text{C}$. This standard contains 18 bacterial strains including five strains of *E. coli* (JM109, B-3008, B-2207, B-766, and B-1109), 2 fungal strains, and 1 archaeal strain in staggered abundances, theoretically ranging from 20.01% to 0.0009% considering the cell number. The list of species, classification, and approximate cell number are listed in table 7. They are referred to as “Gut standards” in the thesis.

Table 7: List of microorganisms with their classification and composition in the gut standard

Species name	Strain Id	Classification	Cell number (%)
<i>Faecalibacterium prausnitzii</i>	AP34BHI	Gram-positive bacteria	14.82
<i>Lactobacillus fermentum</i>	B-1840	Gram-positive bacteria	9.71
<i>Clostridioides difficile</i>	P4D3A1	Gram-positive bacteria	1.10
<i>Clostridium perfringens</i>	OB21 TSA 19	Gram positive bacteria	0.00009
<i>Enterococcus faecalis</i>	IP101412 AER FAA 2	Gram-positive bacteria	0.0011
<i>Roseburia hominis</i>	OB EAV1 11 DCM	Gram invariable bacteria	12.47
<i>Veillonella rogosae</i>	AC2811 AN NA 2	Gram negative bacteria	20.01
<i>Bacteroides fragilis</i>	OB EAV1 11 D6 FAA	Gram-negative bacteria	8.36
<i>Prevotella corporis</i>	OB21 FMU 4	Gram-negative bacteria	6.28
<i>Bifidobacterium adolescentis</i>	LMG 10502	Gram-positive bacteria	8.86
<i>Fusobacterium nucleatum</i>	2/1/50A	Gram-negative bacteria	7.56
<i>Akkermansia muciniphila</i>	OB21 FAA NB 28	Gram-negative bacteria	1.62
<i>Escherichia coli</i>	JM109	Gram-negative bacteria	1.83
<i>Escherichia coli</i>	B-3008	Gram-negative bacteria	1.82
<i>Escherichia coli</i>	B-2207	Gram-negative bacteria	1.65
<i>Escherichia coli</i>	B-766	Gram-negative bacteria	1.66
<i>Escherichia coli</i>	B-1109	Gram-negative bacteria	1.77
<i>Salmonella enterica</i>	B-4212	Gram negative bacteria	0.0065
<i>Candida albicans</i>	IHEM 3108	Fungi	0.16
<i>Saccharomyces cerevisiae</i>	Y-567	Fungi	0.16
<i>Methanobrevibacter smithii</i>	DSM 861	Archaea	0.17

For all the experiments shown in this chapter, approximately 3×10^8 cells per replicate were considered. Upon receiving the cell from the vendor, the cells were aliquoted and stored at $-80\text{ }^{\circ}\text{C}$ until use.

6.2.2. Proteomics analysis

Sample processing was performed by ChipFilter WF5 with sample introduction using a piston syringe and syringe pump as described in Chapter 2. After peptide elution by using a trapping column, the nucleic acid recovery was done as described in Chapter 5.

Peptides were analysed by nanoLC-MS/MS in data-dependent acquisition (DDA) high-energy c-trap dissociation (HCD) mode using an RSLCnano UltiMate™ 3000 System coupled to a nanoESI Q- Q-Exactive HF mass spectrometer (Thermo Fisher Scientific).

The standard gut microbiota was analysed with a Q-Exactive HF mass spectrometer operated in nanoESI (1.6 kV). Full MS survey scans were recorded over the m/z range of 375-1500 with a resolution of 60 000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 60 ms. Up to 20 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 28%, with precursor isolation window at 2 m/z, resolution of 15 000, AGC value of 1E5 with a maximum injection time of 60 ms. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 sec.

For the data analysis, the Sequest HT search engine was used against a home-made database consisting of the protein sequences obtained from UniProtKB (1,133,353 sequences containing 506,763,197 residues) developed specifically for the species comprising the standard mix. This database includes 21 strains from the gut standard-*Faecalibacterium prausnitzii*, *Roseburia hominis*, *Bifidobacterium adolescentis*, *Lactobacillus fermentum*, *Clostridioides difficile*, *Methanobrevibacter smithii*, *Enterococcus faecalis*, *Clostridium perfringens*, *Veillonella rogosae*, *Bacteroides fragilis*, *Prevotella corporis*, *Fusobacterium nucleatum*, *Akkermansia muciniphila*, *Salmonella enterica*, *Escherichia coli* (JM109), *Escherichia coli* (B-3008), *Escherichia coli* (B-2207), *Escherichia coli* (B-766), *Escherichia coli* (B-1109), *Candida albicans* and *Saccharomyces cerevisiae*. The database search was performed with the following parameters: MS and MS/MS mass tolerance 10 ppm and 0.02 Da respectively, trypsin specificity with up to 2 miscleavages, partial Carbamidomethylation (C), Deamidation (NQ) and Oxidation (M). Proteins with at least one high-confidence peptide and six amino acids were validated. Target FDR was set at 0.01. The experiments were performed in three replicates.

6.2.3. Nucleic acid analysis

To compare the amount of nucleic acid isolated by ChipFilter workflow as compared to a commercial method, a commercial kit (ZymoResearch) was used to isolate DNA from the same number of cells used during ChipFilter experiments. For DNA isolation using the kit method, the manufacturer's guidelines were followed.

ZymoBIOMICS DNA Microprep kit protocol

<https://files.zymoresearch.com/protocols/ d4301 d4305 zymbiomics dna microprep kit.pdf>

DNA was quantified using Nanodrop (Thermo Fisher Scientific). To check the integrity of the isolated nucleic acids, DNA was resolved by 1% agarose gel electrophoresis as previously explained in chapter 4.

DNA samples were sent to the Plateforme de Genotypage/Séquençage of the Paris Brain Institute (ICM; Hôpital de la Pitié-Salpêtrière; CNRS UMR 7225 - Inserm U 1127) performed cDNA library according to Illumina sequencing requirements, using the standard protocol. At this date, NGS sequencing has not been realized.

6.3. ChipFilter allows the identification of proteins from most of the species of the gut standard

The proteomic analysis of the gut standard allowed the identification of 2521 ± 375 proteins. At least two or more proteins were identified for 16 out of 17 species considered, whereas only one peptide assigned to the archaeal *M. smithii* was identified (twice out of the three replicates) (figure 56). A total of 6099 ± 1140 peptides, and $18\,339 \pm 1535$ PSMs were identified (table 8).

These results confirm those obtained with the mixed analysis presented in Chapters 2 and 3, showing that microorganisms of different groups can be lysed and prepared in the ChipFilter. These results are also in line with the sensitivity obtained since proteins from cells present in hundreds or thousands of copies in the mix were identified.

Table 8: Number of proteins, peptides, and PSMs identified for Gut standard by ChipFilter WF5

Species Name	Proteins	Peptides	PSMs
<i>Faecalibacterium prausnitzii</i>	408 ± 78	1130 ± 282	3706 ± 395
<i>Lactobacillus fermentum</i>	291 ± 30	760 ± 92	2556 ± 135
<i>Clostridioides difficile</i>	68 ± 3	131 ± 8	332 ± 19
<i>Clostridium perfringens</i>	5 ± 2	5 ± 2	10 ± 3
<i>Enterococcus faecalis</i>	2 ± 0	3 ± 1	10 ± 1
<i>Roseburia hominis</i>	272 ± 46	781 ± 160	2525 ± 227
<i>Veillonella rogosae</i>	167 ± 21	413 ± 73	1269 ± 100
<i>Bacteroides fragilis</i>	220 ± 9	490 ± 18	1823 ± 25
<i>Prevotella corporis</i>	91 ± 6	175 ± 15	431 ± 21
<i>Bifidobacterium adolescentis</i>	113 ± 20	221 ± 54	510 ± 79
<i>Fusobacterium nucleatum</i>	164 ± 28	456 ± 113	1456 ± 162
<i>Akkermansia muciniphila</i>	21 ± 6	28 ± 6	47 ± 7
<i>Escherichia coli</i>	342 ± 74	710 ± 209	1564 ± 274
<i>Salmonella enterica</i>	23 ± 4	61 ± 16	142 ± 22
<i>Candida albicans</i>	91 ± 19	184 ± 33	464 ± 39
<i>Saccharomyces cerevisiae</i>	242 ± 38	551 ± 86	1493 ± 110
<i>Methanobrevibacter smithii</i>	1 ± 0	1 ± 0	1 ± 1

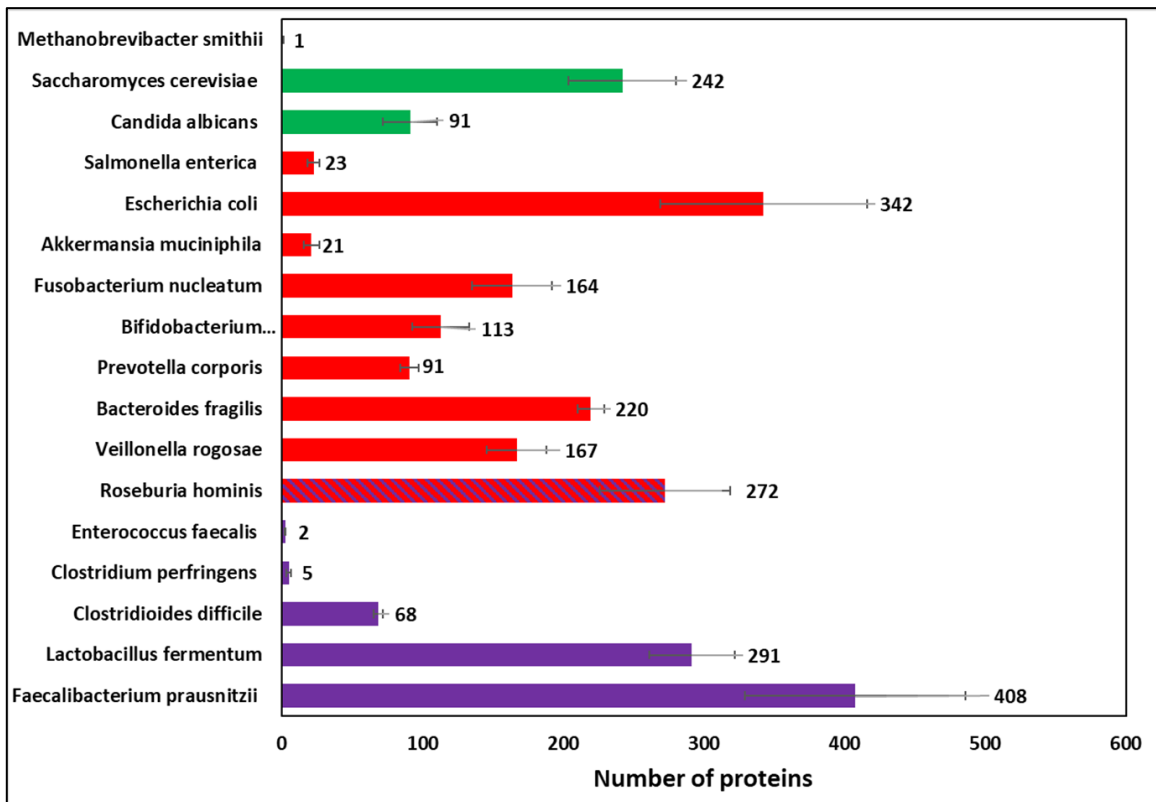


Figure 56: Proteins identified for different species present in the Gut standard by ChipFilter WF5

The colour represents a group that the species belongs like violet for Gram-positive bacteria, red for Gram-negative bacteria, green for fungi, blue for archaea and. *R. hominis* which is Gram invariable is represented in two colours. The five *E. coli* strains in the mixture are grouped under a single species name.

6.4. Evaluation of the effectiveness of the preparation according to the microbial groups

In this section, I will discuss the results obtained for each group of microorganisms i.e., Gram-positive bacteria, Gram-negative bacteria, fungi and archaea.

Gram-positive bacteria

About 34.5% of the total theoretical biomass consists of Gram-positive bacteria (without including *R. hominis* (12.5%)) that including *C. perfringens*, *L. fermentum*, *C. difficile*, *E. faecalis*, *F. prausnitzii*, *B. adolescentis*. *C. perfringens* is a pathogenic bacterium that is found in the human gut and environment (soil or marine sediments) (Ryan & Ray, 2004). *L. fermentum* is a probiotic bacterium that is found in several food products

(Golden et al., 2005). *C. difficile* are pathogenic bacterium associated with diarrhoea and colon cancer (Drewes et al., 2022). *E. faecalis* is a pathogen colonizing the human gut and is associated with infections and chronic inflammation of the intestine (Ryan & Ray, 2004). *F. prausnitzii* is one of the most common gut bacteria and is helpful in the digestion of dietary fibers (Benus et al., 2010). *B. adolescentis* is a commonly identified bacteria after birth in the human gut (Milani et al., 2017). An important development made to the ChipFilter workflow to lyse the peptidoglycan cell wall of the Gram-positive bacteria was to include lysozyme treatment. This procedure was included after the underrepresentation of protein identification in *B. adolescentis* was observed. Except for *C. perfringens* and *E. faecalis* which are present in very low amounts of 360 (0.00009%) and 4400 (0.0011%) cells, the protein identified for other Gram-positive species was over fifty proteins. On the contrary, the ChipFilter workflow has the advantage to detect species-specific peptides from as low as a few hundred cells. Thus, Gram-positive bacteria can be processed and peptides can be identified by ChipFilter workflow.

Gram-negative bacteria

The largest of the total theoretical biomass corresponds to Gram-negative bacteria consisting of 52.57% (without including *R. hominis* (12.5%)) of the total. The list includes *R. hominis*, *V. rogosae*, *B. fragilis*, *P. corporis*, *F. nucleatum*, *A. muciniphila*, *S. enterica*, and five strains of *E. coli* (JM109, B-3008, B-2207, B-766, B-1109). *R. hominis* is Gram invariable and is known to play important roles in gut-related diseases like IBD and Crohn's disease (Nie et al., 2021). *V. rogosae* are lactate fermenting bacteria found in the human gut and oral tract (Zhou et al., 2021). *B. fragilis* is a commensal bacteria found normally in the gut but can be infectious if spread to the bloodstream (Wexler, 2007). *P. corporis* is found in the oral, vaginal, and gut microbiome of humans and are associated with diseases of the respiratory tract (Könönen & Gursoy, 2022). *F. nucleatum* is a commensal bacterium in the oral microbiome and is detected in the gut as well (Signat et al., 2011). *A. muciniphila* is a mucin-degrading bacterium found in the gut (Derrien et al., 2004). *S. enterica* is found in the human gut and often is associated with pathogenesis (Eng et al., 2015). *E. coli* are well-known bacteria in the human gut that demonstrate a symbiotic relationship with the host but, certain serotypes are known to cause food poisoning (Wiebelhaus et al., 2021; Bhattacharyya et al., 2016). Most of the Gram-negative bacteria are well represented in terms of protein identification of more than fifty proteins except *S. enterica* (23 ± 4 proteins) and *A. muciniphila* (21 ± 6 proteins). In the case of *A. muciniphila*, recent reports (Lee et al., 2022) identify 841 proteins that cover only 40% of the protein-coding genes. They report that there is a lack of a database covering functional proteins for the species and this could be reasoned for lower identification. *S. enterica* cell composition in the mix was only $2.6 * 10^4$ cells (0.0065%)

and maybe this might have affected the identification. Overall, the ChipFilter workflow was successful in cell lysis and protein identification of Gram-negative bacteria.

Fungi

Two fungal species *C. albicans* and *S. cerevisiae* were included in the mixture. *C. albicans* is an opportunistic pathogen that is found in the human gut and also outside the environment (biofilms) (Gow & Yadav, 2017; Kumamoto, 2002). *S. cerevisiae* is a yeast that is commonly used in the food industry for fermentation and thereby being identified in the human gut (Parapouli et al., 2020). In the present study, 91 ± 19 proteins from *C. albicans* and 242 ± 38 proteins from *S. cerevisiae* were identified. This corresponds to approximately 14% of the total identified proteins. The cell mixture however contained only 0.32 % of fungal cells which approximately corresponds to $6.4 * 10^5$ cells for each species. In the previous mixture (mix 2) which contains approximately $1 * 10^4$ *S. cerevisiae* cells, more than 750 proteins were obtained. It is not clear why the identification was lowered for the same species when a much more complex mixture with a higher individual cell number was used. One reason could be due to the limitation in the instrumentation that was operated in DDA mode with Top20 ions. Performing in DIA mode was shown to identify more than 2000 proteins from *C. albicans* from 50 μg of protein lysate (Amador-García et al., 2021).

Archaea

M. smithii is a species of methanogenic archaea that is largely identified in the human gut. Despite being present at a cell number of $6.8 * 10^5$ cells (0.17%), the peptides of this species were not identified in two replicates (Samuel et al., 2007). It was thus considered to be not identified by the ChipFilter workflow. A possible explanation will be that the cell lysis method used in the ChipFilter was not designed to lyse the archaeal cell wall that is ether-linked lipids, unlike the bacterial cell wall that has ester-linked lipids. Limited proteomic studies have been done on *M. smithii* archaeon and thus the direct comparison of workflows was difficult. Hence, studies from another model archaeon known as *Haloferax volcanii* is highlighted. In this study, (Jevtić et al., 2019) 2244 proteins were identified by SWATH-MS from 50 μg protein lysate and the sample preparation involved anionic detergent-based cell lysis (sodium taurodeoxycholate, 0.006% final concentration), followed by nuclease treatment and protein precipitation with acetone. It will be of interest in the future to use optimal cell lysis conditions in ChipFilter for archaeal samples.

6.5. Biomass estimation by proteomics

Proteins contribute the maximum to the species' biomass and biomass estimations can be much more accurate to determine the function of the species within a community. Hence, biomass estimation methods are critical to microbiome studies. Cell number estimation which is an indirect method to assess the species participation within a community can be currently done by shotgun DNA sequencing or targeted sequencing of the 16S or 18S rRNA regions. Towards this several reports have been made to estimate biomass by metaproteomic approaches and in this study, I use the principles provided by Kleiner et al., (Kleiner et al., 2017).

The principles govern the choices to be made to identify the protein biomass contribution in metaproteomics to overcome the protein interference problem (matching of the same peptide sequence to multiple different proteins which may lead to ambiguity). The important points (Kleiner et al., 2017), which are relevant in the context of the gut standard experiment, are listed below:

1. For Kleiner et al., a single 260 min LC-MS/MS runs and data acquisition, on a Qexactive plus mass spectrometer with 2 µg of the peptide was sufficient to distinguish 25-26 species from a list of 30 uneven microbial species (bacteria, archaea, virus, and algae) biomass.
2. Considerations to improve sensitivity and specificity in protein identification by increasing the false discovery rate for target-decoy search to 5%, identifying proteins with a minimum of two unique peptides, and using tools like Fido for proteome discoverer.
3. The availability of accurate protein sequence databases having high similarity to the actual protein was necessary to have a good estimation of biomass.
4. The PSM-based counting and summing method for label-free quantification proved to be robust for estimating abundances for the uneven microbial community (different cell numbers and total protein amount between species).
5. Metaproteomics proved to be more accurate for biomass estimates for every species (bacteria, algae, archaea, and phage) than other sequencing methods for equal protein amount or uneven cell number communities.

According to these considerations and our methodological setup, I adapted my protocol for biomass estimation of gut standard as follow:

1. For LC-MS/MS, 120 minutes gradient for a Qexactive HF mass spectrometer was used. Since there was no observed crowding of ions in TopN spacing, as seen in figure 57, it was not modified for the gut standard sample in this study.
2. I use quantification with 5% FDR and at least 1 unique peptide.

3. The protein database was constructed with high specificity using the UniProtKB.
4. The PSM counting was performed to quantify biomass from uneven cell number community in gut standard.
5. The results are compared with the cell number, 16S and 18S, and genome copy number information provided by the manufacturer (figure 58).

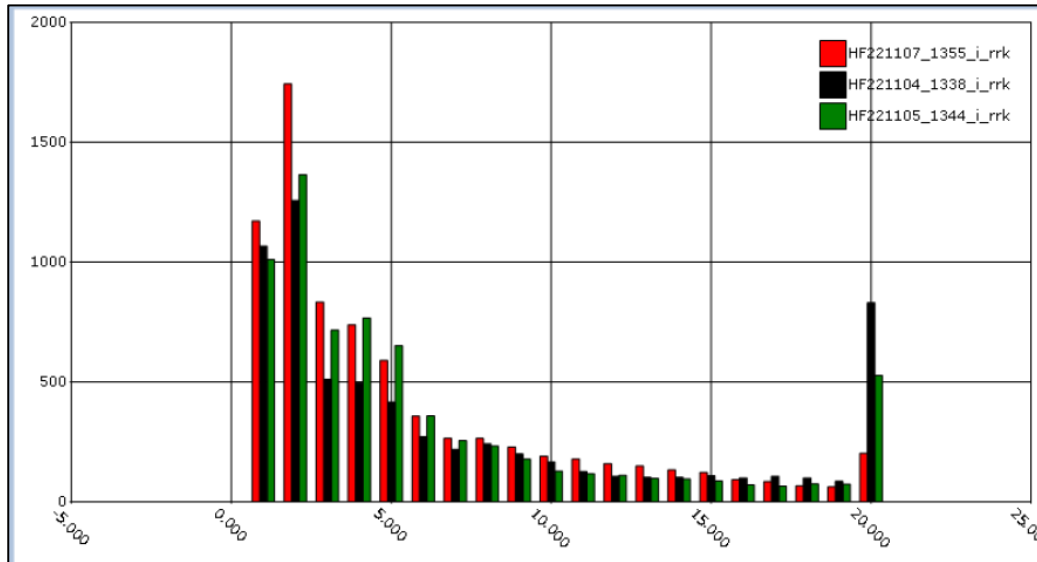


Figure 57: TopN spacing distribution of precursor ions selected for fragmentation by MS/MS as seen in RawMeat version 2.1 for the three replicates of gut standard

The results obtained by estimating the biomass contribution close to the estimates provided by the manufacturer in terms of the cell number or targeted sequencing are explained. For species, *R. hominis*, *E. coli*, *F. nucleatum*, and *B. fragilis* the relative accuracy seen between the ChipFilter biomass estimate and cell number/genome copies were with low fold change (-0.02 to 0.19), implying that they were accurate. While for the eukaryotic *C. albicans* (fold change -0.19) and *S. cerevisiae* (fold change 0.28), the ChipFilter estimate matches well with 18S sequencing results than 16S or cell number results. Finally, *C. difficile* (fold change -0.31) and *F. prausnitzii* (fold change 0.15) has a closer estimate to 16S sequencing than the cell number estimated. The indifference in estimation can be greatly seen for bacterial species that have either the highest protein identification like *L. fermentum* or with lower protein identification species like *V. rogosae*, *B. adolescentis*, *P. corporis*, *S. enterica*, *A. muciniphila*, *C. perfringens* and *E. faecalis*. *M. smithii* identified only PSM and thus cannot be considered for this comparison. The observed difference in estimates can be reasoned due to a lack of sufficient protein sequence databases (like in the case of *A. muciniphila*), limited data acquisition by LC-MS/MS (lower gradient time of 120 minutes was used than mentioned 260 minutes), and methodological improvements that need to be made to ChipFilter

workflow to obtain information of the representative species. The results show promise for future studies as the metaproteomic samples prepared by ChipFilter workflow can aid in biomass estimation too.

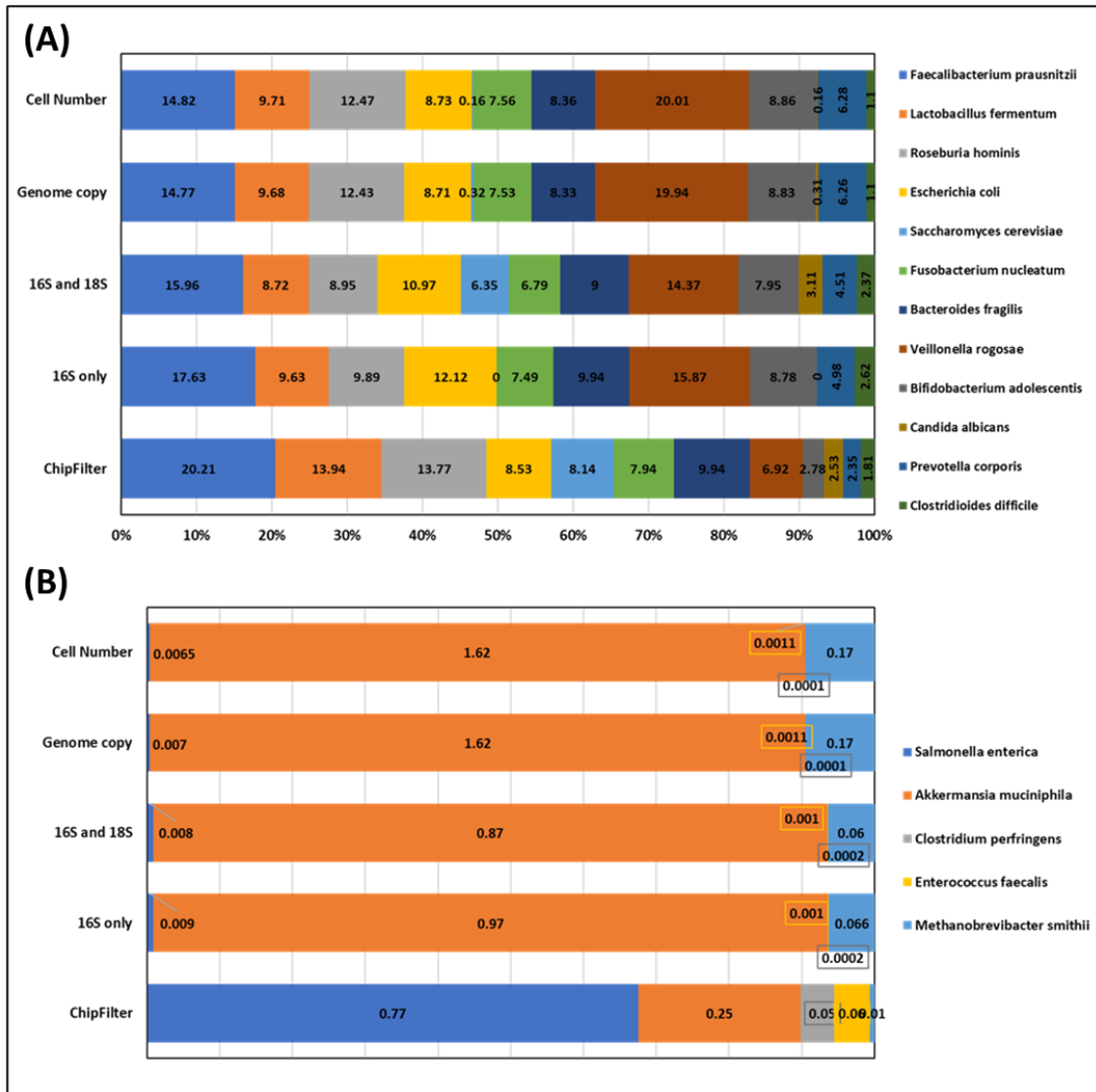


Figure 58: Comparison of biomass estimate done by ChipFilter-based metaproteomics with the next-generation sequencing methods and cell number for more than 1% abundance (A) and less than 1% abundance (B). Data for cell number, genome copy, 16S, and 18S were provided by the manufacturer.

6.6. Nucleic acid retrieval

Nucleic acids were retrieved as per workflow 5 and quantified using nanodrop (Thermo Scientific). The nucleic acid elution was performed twice and it was identified that the nucleic acid was present only in the first elution. Nucleic acid concentrations of 39.4 ± 22.4 ng/ μ l (DNA mode operation) and 21.3 ± 13.5 ng/ μ l (RNA mode operation) for 30 μ l of elution buffer were obtained. The high standard error between the replicates and low purity (260/280 measurement for ChipFilter by WF5 was 1.50, while for the kit method, it was 1.90) were a cause for concern. Further research is necessary to overcome this bias.

To compare the amount of nucleic acid isolated by ChipFilter workflow as compared to a commercial method, I used a commercial kit (ZymoResearch) to isolate DNA from the same number of cells used during ChipFilter experiments. The DNA isolated by this method was 11.7 ± 0.17 ng/ μ l for 30 μ l of elution buffer. This method also produced highly purified DNA without protein contamination (260/280 measurement for the kit method was 1.90).

To check the integrity of the isolated nucleic acids by ChipFilter workflow, DNA was resolved by 1% agarose gel electrophoresis to figure 59. The fragmented DNA or RNA can be seen at the bottom, with a size of less than 200 bp. Also, a dense band that is not migrated is retained in the well (figure 59A). This suggests that the nucleic acid obtained by the ChipFilter method has a mixture of fragmented and cell component-bound nucleic acid along with the desired nucleic acids. Purification of the nucleic acid obtained after ChipFilter workflow may be added to enhance the quality, if necessary for further applications.

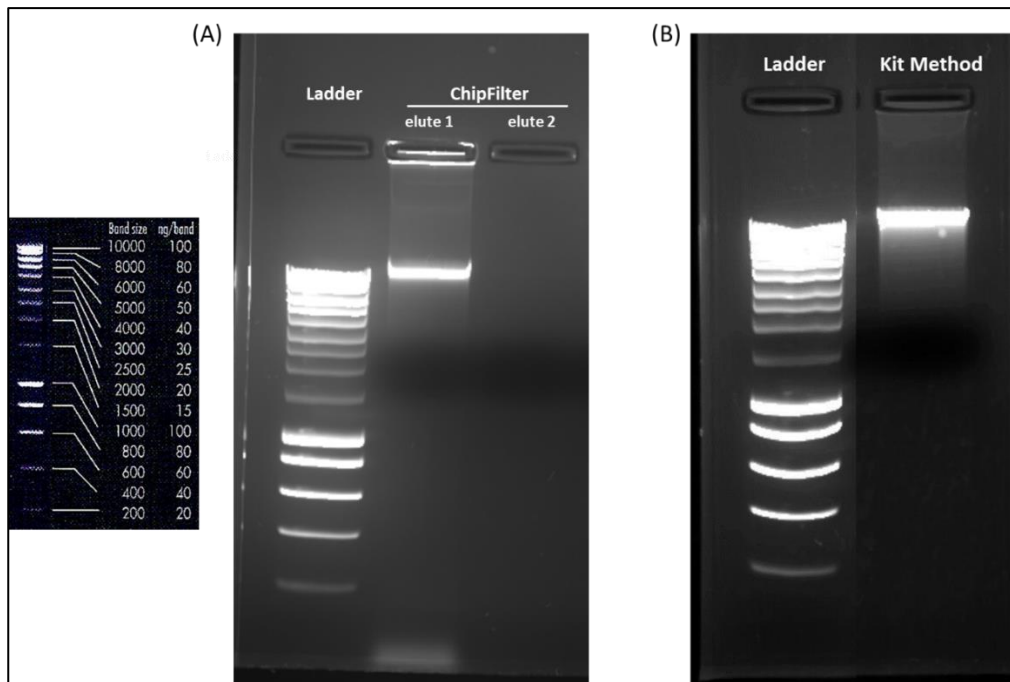


Figure 59: Nucleic acid integrity obtained after ChipFilter WF5 (A) and DNA isolation kit (B).

Finally, the ChipFilter isolated nucleic acids were subjected to shot-gun sequencing along with the DNA isolated by kit method to estimate the sequencing efficiency between the two methods. The initial results obtained by the DNA size and quantity estimation by TapeStation (Agilent Technologies), showed that two out of the three samples prepared by ChipFilter workflow have passed the requirements. In comparison to the kit method, the quantity was lower (figure 60). NGS is yet to be completed and thus the results will not be shown here.

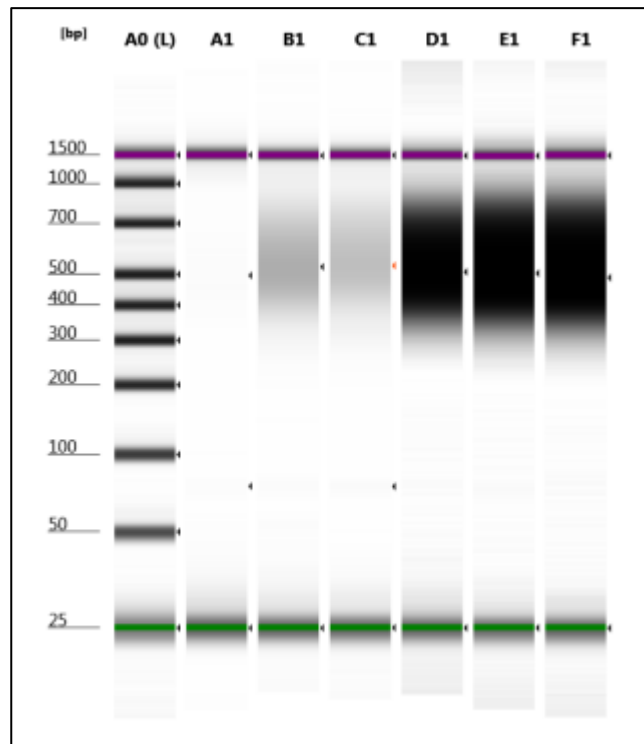


Figure 60: Nucleic acid integrity obtained during library generation using TapeStation after ChipFilter WF5 (A1-C1) and DNA isolation kit (D1-F1). A0 (L) refers to the ladder

6.7. Conclusions

A standard uneven cell mixture of 21 species belonging to Gram-positive bacteria, Gram-negative bacteria, fungi, and archaea that are commonly found in the human gut was taken and processed by ChipFilter workflow. On the same samples, preparation for proteomic followed by nucleic acid retrieval was made.

Proteins were identified for all the fungal and bacterial cells but not for archaeal cells. Biomass estimation using the PSMs was made to determine the accuracy between the ChipFilter metaproteomics workflow and nucleic acid sequencing, which showed good accuracy for more than half of the identified species.

The nucleic acid retrieved was characterized for quality and quantity by gel electrophoresis and nanodrop estimation. Samples could be prepared for NGS sequencing.

In conclusion, the preparation of proteins and nucleic acid obtained through ChipFilter workflow can be a very novel strategy to steer the multi-omics sample preparation. Despite the advantage and the demonstration of the viability of this work, several key places need to be addressed to fully utilize the ChipFilter workflow for multi-

omics sample preparation. On the nucleic acid level, the quality of the isolated nucleic acids can be enhanced to obtain sequencing depth. The method can also be optimized for RNA recovery. At the protein level, cell lysis of other microbial species (like archaea) apart from fungi and bacteria needs to be established by ChipFilter workflow. It could be adapted to the sample to be analysed. Finally, the NGS sequencing results will allow the integration of genomic and proteomic data for meta-omics analysis.

7. ChipFilter adapted for aerosol-borne microbial sample collection

7.1. Introduction

The earlier chapters discussed the development of ChipFilter for metaproteomics and further isolate nucleic acids to perform nucleic acid-based sequencing studies to complement and gain deeper insights into the microbiome. This leads to the progress of ChipFilter from a proteomic sample preparation device to a multi-omics sample preparation device. The models used were microbial suspensions, a commercial gut standard microbiome, and a whole sample of murine gut sample.

However, other microbial environments may be interesting to study with a multi-omics approach. The Covid crisis has reminded us of the importance of studying aerosols, which contain microorganisms, viruses, and free proteins. Their impact is of crucial importance. As introduced before cloud offers a great environment for microbiome propagation and the bacteria identified have shown abilities to form rain clouds by initiating ice nucleation. To study these aerosols, in which cell concentrations are very low, it would be interesting to collect the samples directly with the ChipFilter. This would reduce the risks of loss of material and contamination by the external environment.

The objective of the work presented in this chapter is to evaluate the feasibility of using the ChipFilter for sample collection (figure 61). Several modifications to the ChipFilter design were tested to allow aerosol sample collection. The technical challenge is to pass a sufficient volume of aerosols through the ChipFilter without modifying its physical characteristics, in particular the sealing between the two chambers. This chapter will introduce the design modifications and show some of the applications of the modifications in actual conditions.

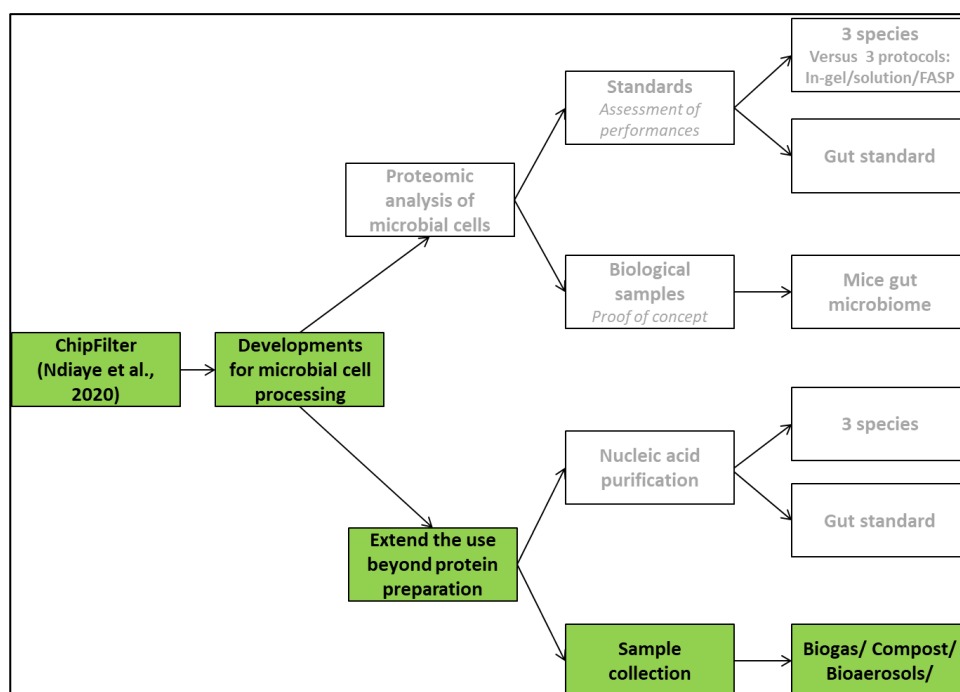


Figure 61: Steps discussed in chapter 7

7.2. ChipFilter directly connected to a pump for sample collection

7.2.1. Experimental design

To test the capacity of the ChipFilter to withstand the high pressure caused during fluid flow and trap the particulates (cells suspended in small droplets), I tested two methods where (A) the atmospheric gases are pumped inside the ChipFilter using a portable air compressor (Xiaomi) and (B) the particles are sucked through the filter of the ChipFilter device using a laboratory vacuum pump. For the air compression method (A), the ChipFilter was connected to the compressor using a PEEK capillary to accommodate the difference in the valve diameter between the instrument and the inlet of the ChipFilter. For the vacuum pump method (B), the outlet of the pump was directly connected to the ChipFilter as the vacuum generated allowed for complete sealing. These experiments were performed to check the stability of the setup (i.e., micropillars supporting the filter in the reaction chamber) to withstand the pressure generated due to the flow of the gases. The setup of the experiment is shown in figure 62.

After sample collection, the device was directly connected to the Fluigent pump system, and the sample was processed with workflow 2 described in Chapter 2. The resulting peptides were analysed by LC-MS/MS as described earlier in section 5.2. Protein search was performed using proteome discoverer 2.4 with the Mascot search engine

(Matrix Science Mascot 2.2.04) with the same conditions mentioned in section 5.2, using the all taxa Swissprot database version release 2022_03 with 568 002 sequences; 205 171 419 residues.

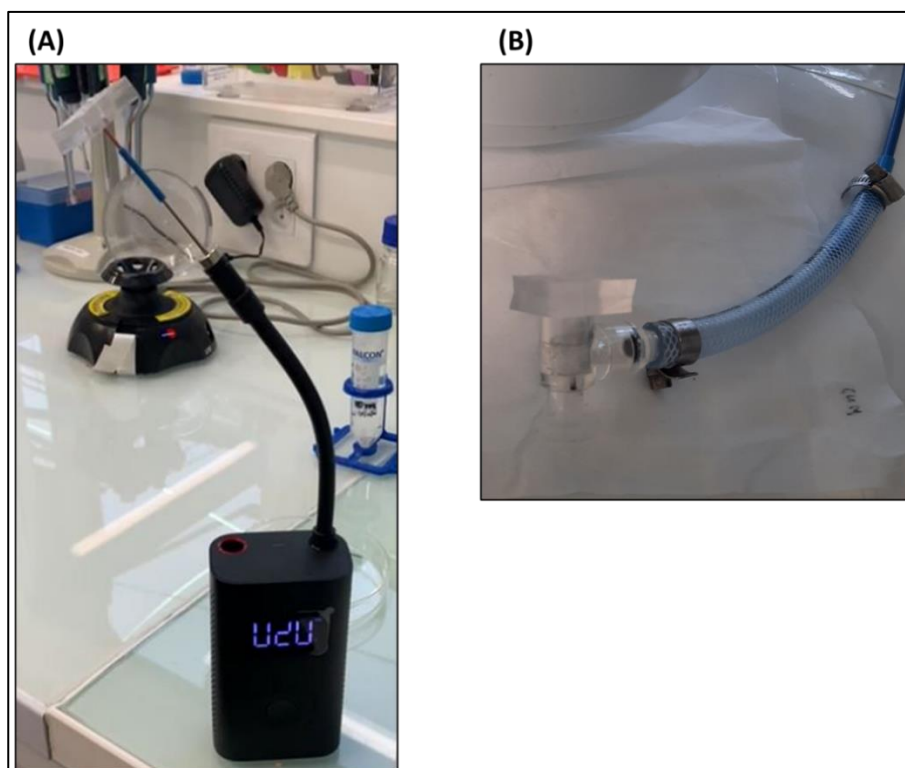


Figure 62: Setups to check the stability of the device for aerosol collection with (A) air compressor and (B) vacuum pump

7.2.2. Application

Bioaerosols

This study aims to test the capacity of the device to withstand the operational conditions necessary for sample collection using ChipFilter.

The preliminary experiments indicate that a pressure of over 1.5 bar (22 psi) severely damaged the ChipFilter, rendering it unusable. It was also observed that the air compression method shown in figure 63A caused easier damage to the ChipFilter than the vacuum pump method (figure 64). The operational pressure has to be maintained under 20 psi (1.37 bar).

Air samples from the laboratory were collected for one hour according to the air compressor method. A very small number of proteins were identified corresponding to

human keratins and common biological contaminants. It is difficult to interpret these results to make a conclusive remark on the identification or compare the two methods in terms of effectiveness because of the lack of controlled environments (which was difficult to design). However, these results suggest that it is possible to trap microorganisms directly on the ChipFilter.

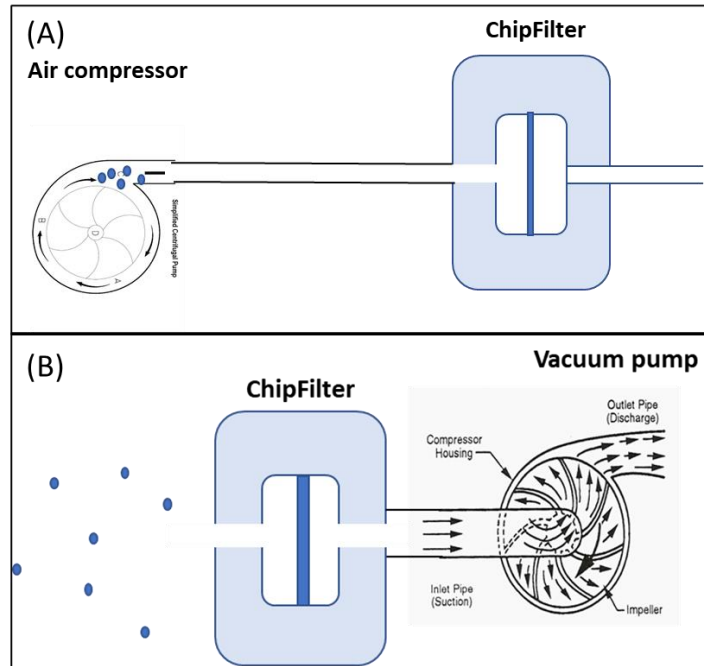


Figure 63: Representation of the setup to test the stability of ChipFilter using air compressor pump (A) and vacuum pump (B)

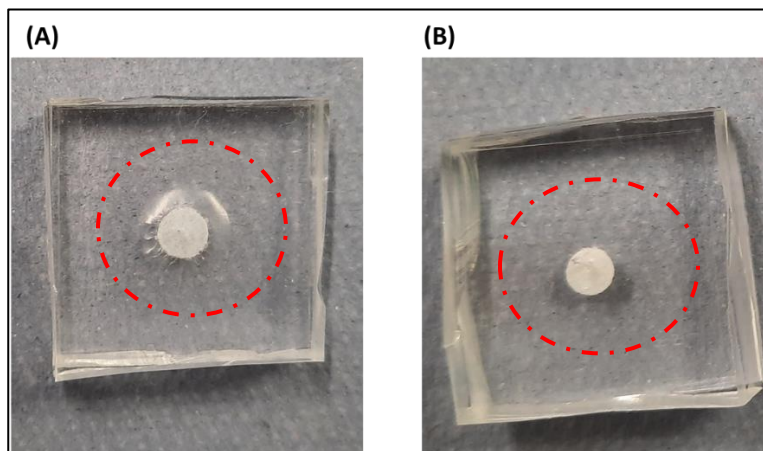


Figure 64: ChipFilter stability analysis shows disruption of the reaction chamber (A) at high injection pressure as compared to control (B)

Composts

To have an aerosol in an environment richer than the air of the laboratory, I analysed the aerosols emitted by domestic compost. The device was connected to a pump and sampling was done for four hours. MS analysis from the samples identified proteins from rats (*Rattus sp.*), dogs (*Canis sp.*), and humans most of which are extracellular in origin like keratins. It was interesting to observe proteins from dogs and rats that have mostly polluted the environment. A statistically significant conclusion cannot be made in this case due to the lack of experimental replicates.

7.3. ChipFilter with 18-gauge needle for sample collection

7.3.1. Experimental design

In this setup, modifications were done to the original design of the ChipFilter by infusing an 18-gauge needle (Terumo) at the entrance and a connector (ufluidic) at the exit (figure 65). The installations of the needle and connector were made without disrupting the original device structure. First, the needle and connector were made to bind with PDMS, it was done by allowing the liquid PDMS to polymerize with the needle and connector. Then, it was fused to the ChipFilter by surface plasma treatment (20 W, 8 sccm O₂ flow, and 0.13 mbar pressure for 1 min).

The flow was created by vacuum mode where the particles at the end of the needle are sucked inside the ChipFilter to exit through the connector, thereby trapping the particles with the ChipFilter.

As previously described, after sample collection, the device was directly connected to the Fluigent pump system, and the sample was processed with workflow 2 described in Chapter 2.

The resulting peptides were analysed by LC-MS/MS as described earlier in section 5.2. Protein search was performed using proteome discoverer 2.4 with SequestTM search engine with the same conditions mentioned in section 5.2, using the all taxa Swissprot database version release 2022_03 with 568 002 sequences; 205 171 419 residues.

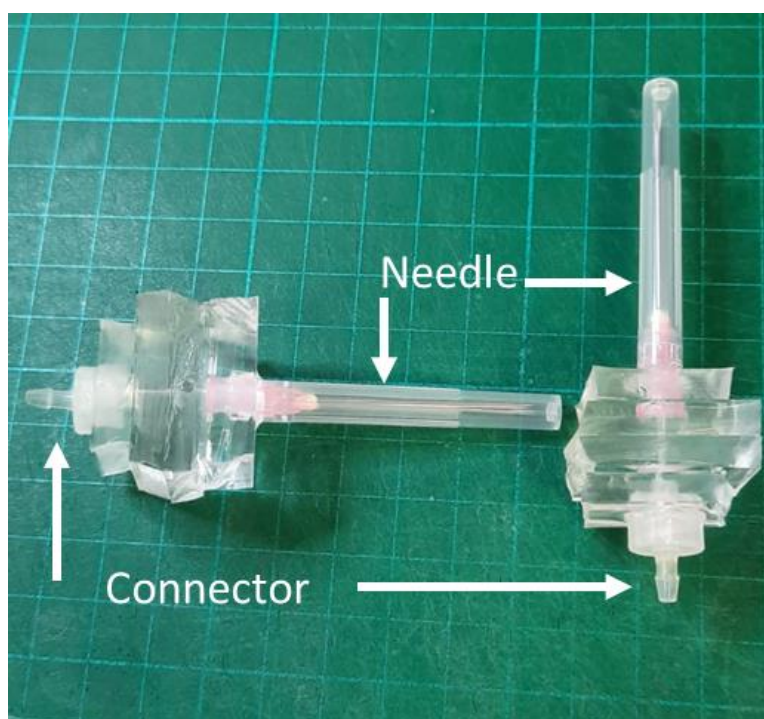


Figure 65: ChipFilter modification to accommodate the connector and needle

This modification to the ChipFilter was done to allow sterile sampling as compared to other methods discussed earlier that identified many contaminants. Also, the needle allowed the introduction of the ChipFilter into the site of sampling such as bioreactors. It can also aid in sampling. The connector at the exit was optional and only used in the case of bioreactor sampling where the released gas was trapped for particles using ChipFilter and the remaining gases were analysed for the gas composition to monitor the reactions inside the chamber.

7.3.2. Applications

Bioreactors

This setup was made to directly allow the sampling of the aerosols from the bioreactors that house microorganisms digesting sludge. This work was realized thanks to a collaboration that I initiated with Dr. Dirk Benndorf's team from Otto-von-Guericke-Universität Magdeburg, Germany. The objective was to identify bacteriophages and bacterial proteins released at different time points during the sludge degradation and methane release process inside a bioreactor. The setup and the operation can be seen in figure 66. Protein identifications cannot be disclosed due to confidentiality with the partnering institute.

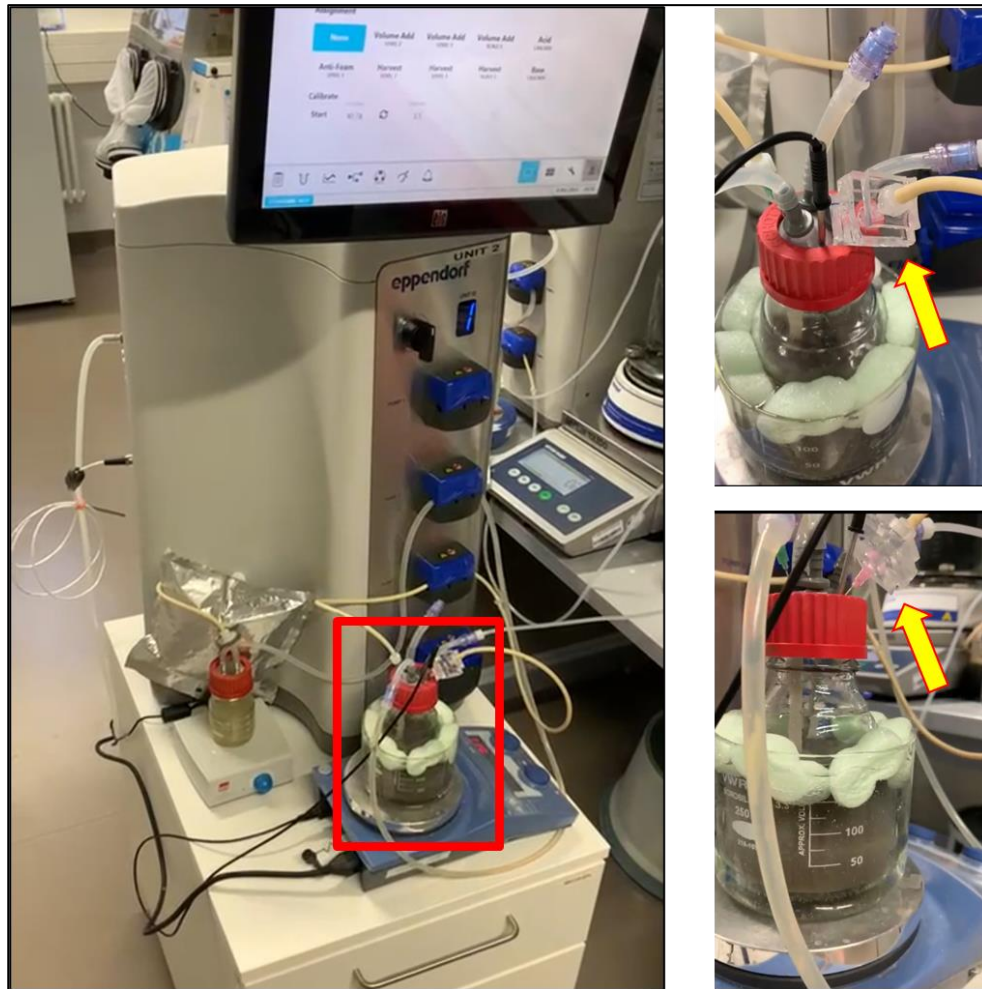


Figure 66: ChipFilter application in bioreactor sampling

Breath analysis

The needle was also tested for blowing breath inside the ChipFilter, and this experiment setup can be used to collect oral and respiratory tract microbiomes. Although, the study was done only as a proof of study, and the test was conducted with the consent of the participant (myself). The identification of microbial proteins was not achieved due to the lack of a reference protein sequence database for this type of microbiome. It can be perceived in the future as the spectra are currently available.

7.4. Conclusions

I have used the ChipFilter for aerosol sample collection, without or with modification using additional components like a syringe. This work showed the stability of the device to withstand air compression and suction over a long time.

Sample collection, processing, and protein identification were possible with this modified device, although it is yet to be fully used for metaproteomics.

There are only a few works on the metaproteomic study of aerosols (Liu et al., 2016; Piovesana et al., 2019). The low cell concentration in aerosols is indeed a major challenge.

There are different ways to approach this challenge. At the pre-analytical level, ChipFilter's ability to perform sample collection and preparation for proteomic and genomic analysis is a major asset. One limitation of this system is the small volume of aerosols that can pass through the chip unless the trapping time is increased. The size of the chip can however be an advantage for the study of particular cases. We have seen that the chip can be interfaced with a bioreactor, in this example the size of the chip allows for the concentration of the cells and is an advantage. We can imagine other examples, such as mounting the chip on a drone to take samples at high altitudes. It should be kept in mind that the microorganisms in aerosols are not the same as in the gut microbiome. For example, there are many more spores. It would therefore be necessary to adapt our lysis buffer.

But to reach a sensitivity allowing to have relevant results for aerosol analysis, it will not be enough to work on the pre-analytical stage. At the analytical level, the development of more and more sensitive machines opens new possibilities. Moreover, new analytical strategies such as DIA have allowed an important gain in sensitivity, which will be essential for the study of aerosols.

Finally, as discussed in the previous chapter the importance of the data processing steps. A major challenge here is to have the right database. In the case of samples such as aerosols, which contain many species whose genome is unknown, this should be constructed with care.

8. Overall conclusions and future perspectives

The thesis work demonstrated the usage of ChipFilter-based workflows for microbial sample preparation to be analysed by LC-MS/MS by bottom-up proteomics for protein identification. Further, the nucleic acids that were retained in the ChipFilter after protein sample preparation was collected to be studied by sequencing or non-sequencing approaches, allowing multi-omics sample processing for the microbiome.

The ChipFilter workflow for whole bacterial (Gram-positive *B. subtilis* and Gram-negative *E. coli*) and fungal (*S. cerevisiae*) cell lysis and protein sample preparation was achieved. This establishes the option of cell lysis with subsequent protein extraction and protein lysis using ChipFilter workflow. All the necessary conditions used for microbial sample processing corresponding to lysis buffer composition, modes of sample introduction and duration for chemical and enzymatic reactions have been well stated. Although bacterial and fungal cells were efficiently processed, the workflow for whole-cell processing is limited to them and future studies can be built around developing cell lysis of archaeal, viral, and microalgal cells. This can be done by incorporating more severe cell lysis buffers or through acidification (with TFA) of the samples after introduction.

As microbial community samples can be more diverse, a mixture of three species with a defined number of cells was used to study the processing efficiency of ChipFilter. At first, a comparative study between two workflows differing in peptide collection after protein lysis was made to explain the logic between their distinction. The first workflow (WF1) allowed the elution of peptides into a trapping column after protein lysis and the second workflow (WF2) allowed the collection of peptides in a vial and introduced separately with an LC syringe. WF1 allows to capture of peptides directly without loss during elution but fails to capture the peptides released during the digestion step. Meanwhile, in WF2 loss can be experienced after elution due to increased steps, but peptides released during digestion can be captured. WF1 allowed complete automation, while WF2 required manual effort to process the peptides (drying, resuspension, and injection) upon collection. A Peptide collection workflow incorporating the advantages of both WF1 and WF2 can be developed in the future. And if this release can be coupled directly to ESI-MS it can rapidly reduce the identification time for targeted proteomics through automated sample processing.

As a next step, the mixed cells were used in two cell densities – mix 1 (three million) and mix 2 (three hundred) to compare the performance of ChipFilter with other well-established workflows in proteomic sample preparation like mFASP, in-gel, and in-solution. For mix 1, the peptide identification was significantly higher for the ChipFilter method than other methods with the distribution of the identification showing a high percentage of exclusive peptides corresponding to the ChipFilter workflow. The nature of

peptide was also different for ChipFilter in terms of mass distribution, amino acid length distribution, more missed cleavage peptides, and reduction of hydrophilic peptides. For mix 2, peptide identification was higher than mFASP and in-gel method. The peptide identification was better and unique for ChipFilter workflow as compared to other studies for microbial sampling at different cell numbers. The study on the repeatability between replicates by label-free quantification of the peptide ion abundance shows lower efficiency for ChipFilter over other methods. Although the ChipFilter has operational advantages, sensitivity, and robustness in identification, the repeatability needs to be improved.

Metaproteomics studies require microbial community-specific protein or gene sequence libraries to convert the spectral information to peptide sequences with search engines. In most cases, proteomic studies are accompanied by genomic sequencing on the same collected sample with different workflows. For the first time, this study demonstrated the potential to isolate nucleic acids sequentially after proteomics sample processing from the same sample. Several settings including the minimum cell number, the effect of proteomic sample processing reagents on the nucleic acid material, and nucleic acid loss and retention mechanisms have been established. Validation of the plasmid isolation was made by PCR amplification of specific regions within the plasmid. Although sequential nucleic acid isolation was proved possible, the quantity and quality of the nucleic acids were not as much as seen from a dedicated isolation procedure. Thus, changes to the processing methods or post-processing for the obtained nucleic acids are proposed. Developing the ChipFilter fabrication with substrates specific to nucleic acid retention can be beneficial to increase the quantity.

To further increase complexity and introduce the multi-omics workflow for sample processing with ChipFilter, a new mixture comprising seventeen microorganisms identified in the human gut was used. Successful identification of two or more peptides corresponding to sixteen of the species was achieved. Further, the biomass abundance estimated by proteomics was compared to that identified by nucleic acid sequencing methods (considered to be the gold standard) provided by the manufacturer, and more than half of the species abundance matched with it thereby establishing a correlation. Also, the nucleic acid post-protein processing was isolated and submitted for sequencing to be compared with the established isolation workflows. In the future, the results will be compared to the proteomics study to establish interesting multi-omics results through ChipFilter sample processing.

As a proof of concept for microbial sample processing with ChipFilter, microbial cells isolated from the murine faeces corresponding to the murine gut microbiome were

studied. Since no available proteome sequence database exist for this type of sample, two approaches involving protein databases constructed from 50 core microbial genus known in murine gut or MAG translated protein sequence was used. Superior confidence in identification was achieved by MAG translated bioinformatic search approach. But due to the lack of protein annotation in MAG based approach, the core set approach was used for the identification of proteins and the matching microbial species. This indicated the dependence of metaproteomics on complementary omics studies such as genomics. A core set of 28 genera was identified as common to all the replicates and the majority of the species identified corresponded to phylum Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria. Further improvements with bioinformatic searches can enable functional studies, metabolic networks, and protein interaction studies.

To further equip the ChipFilter for microbiome analysis and automation, the collection of aerosol samples was tested without and with modifying the design of the ChipFilter. The integrity of the structure was tested in both cases after sample collection. More developments and experimentation need to be performed to enable together sample collection, processing, and identification. The current study is preliminary to show the viability as several microfluidic devices have been developed specifically to collect aerosols with a design, unlike the ChipFilter.

Despite the many uses and advantages of ChipFilter workflow, it is not fully developed to perform everything necessary for metaproteomic sample preparation. Here are some of the limitations at present seen in ChipFilter with suggestions that can be taken for further development of the ChipFilter.

- Is a microfluidic device adapted to study the microbiome?

The ChipFilter, a microfluidic device helps to manipulate very small volumes allowing efficient catalysis, but not all microbiome samples necessarily have to be handled in this manner. Thus, it can be possible to use methods beyond microfluidics to process and choose to multiplex over automation, as ChipFilter has not been currently exploited for multiplexed sample processing. Other limitations include, the processing of microbial taxa like archaea, algae, or viruses. They have not been currently processed with workflows discussed in the study. Thus, need to be standardized and maybe processed externally for cell lysis. In conclusion, ChipFilter can be very useful to process low sample numbers or volume samples such as biofluids or single cells, the need to process bulk samples may not need ChipFilter. Additionally, the current capacity for ChipFilter-based cell lysis is limited to bacterial or fungal cells in standard cultures or dispersed samples. Other samples like murine faeces microbiome that has contact with biosolids or fibres, archaeal and algal cells will require additional external cell lysis steps.

- What kind of microbiome sample can benefit from ChipFilter?

As a proof of concept, it was demonstrated to use ChipFilter for murine faecal microbiome processing, but the real intention of the ChipFilter was to utilize aerosol samples from the atmosphere and breath. Due to the lack of model systems and a protein sequence database, the project was not achieved. The logic behind this recommendation is that these samples are difficult to collect and are already being collected by trapping with mesh filters and transferred to another system for processing. ChipFilter can offer the advantage of trapping and processing in the same system and can reduce time, sample loss, and contamination. Thus, I would recommend the usage of the ChipFilter for samples that are highly dispersed and require intensive collection steps.

- What are the challenges with ChipFilter-based multi-omics study?

At first, the complete NGS workflow was not studied for ChipFilter-isolated nucleic acid due to limitations in time. But the results will help to decide if the workflow already can process samples for multi-omics or not. In case the method may fail or underperform as compared to the kit method used for comparison, the possible reasons can be due to poor quality nucleic acids with bound chemicals or due to loss encountered due to the inability of the filter used in the ChipFilter. In that case, the development of ChipFilter with membranes capable of retaining nucleic acids or surface treatments can enable nucleic acid retention. In the case of reducing contaminants, purifying the obtained nucleic acids can greatly enable effective multi-omics sample preparation.

- What other uses can ChipFilter have?

ChipFilter was shown to perform microbial collection, cell lysis, protein-to-peptide processing, and recovery of nucleic acids from the same sample. This is a complete proteogenomic workflow for sample preparation. Although, this has to be fully validated for different types of model microbiomes to be necessarily used for the microbiome. Further, post-translational studies using ChipFilter-based sample preparation for metaproteomics can be very attractive. ChipFilter workflow for ocSILAC to study redox omics was already well established, it must be further translated to metaproteomics samples. Additionally, the ChipFilter can be used to perform targeted proteomic sample preparation for rapid and automated screening of pathogens from low-volume samples. Finally, sample preparation for RNA or metabolites can be also accommodated in the ChipFilter by modifying the workflow or the design.

- What are the general limitations of the study?

In addition to the discussed ChipFilter-specific limitations, several broad challenges in metaproteomics may influence the consideration of the ChipFilter. If we consider metaproteomics in three broad steps- pre-analytical, analytical, and post-analytical, only one-third of the requirements have been successfully met in this study and the remaining needs to be addressed. Some of the studies discussed in the project like murine gut microbiome studies have been incomplete during the post-analytical steps. Further, multi-omics studies can be data intensive and will require well-curated pipelines to fully utilize and appreciate the pre-analytical process. Finally, the kinetics of cell lysis, proteolysis, and nucleic acid retention inside ChipFilter needs to be completely understood to effectively use the system. Microscopy-based monitoring of labelled proteins or cells can be one way to perform this procedure.

Overall, the study aims to adapt the ChipFilter for metaproteomic studies and cater to different experimental needs like cell lysis, protein processing, nucleic acid retrieval, and even sample collection. The ambitious goal of using a single microfluidic device for wide-ranging applications is successfully met in the present work that will allow future metaproteomic studies to use the ChipFilter-based sample processing.

Annex

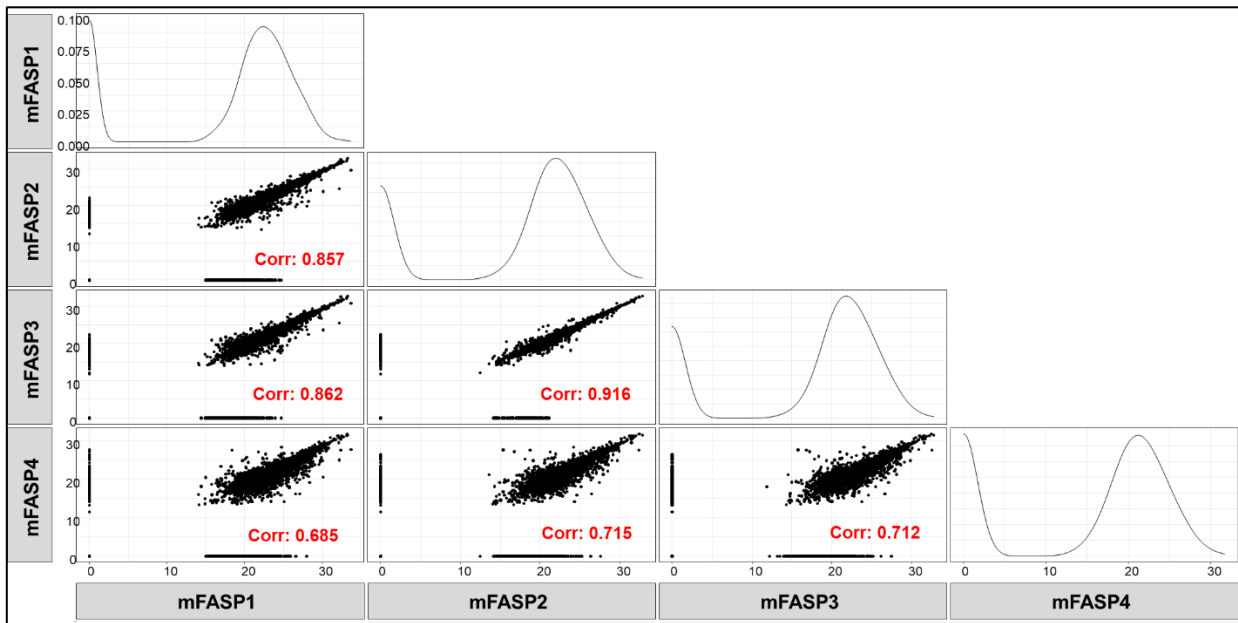
Annex 1 – Description of the ChipFilter Workflows

	Cell loading	Cell washing	Cell lysis	Reduction/alkylation	Washing	Connection	Proteolysis	Peptide Elution	LC	Nucleotides Elution
Workflow 1 (WF 1)	cells loaded into the device	20 µl of 50 mM ABC	lysis buffer (45 µl at 2 µl/min and 30 µl at 1 µl/min)	20 mM DTT in lysis buffer (45 µl at the rate of 2 µl/min and 30 µl at the rate of 1 µl/min) 50 mM IAA in lysis buffer (45 µl at 2 µl/min and 30 µl 1 µl/min)	50 mM ABC (45 µl at the rate of 2 µl/min and 30 µl at the rate of 1 µl/min)	Direct connection to precolumn of the LC system	<u>Step 1:</u> 2µg of trypsin in 50 mM ABC (20 µl at the rate of 2 µl/min) <u>Step 2:</u> 50mM ABC buffer (20 µl at the rate of 2µl/min and 50 µl at the rate of 0.5 µl/min)	50 mM ABC buffer of 15 µl at the rate of 1 µl/min and 30 µl at the rate of 2 µl/min	connect the precolumn by changing the mode to inject to the LC, Transfer the peptides to the trapping column	
Workflow 2 (WF 2)	Same as WF 1					Connect to a vial	Same as WF 1		Dry the vial content in a vacuum drier, reconstitute in 0.1 % TFA	

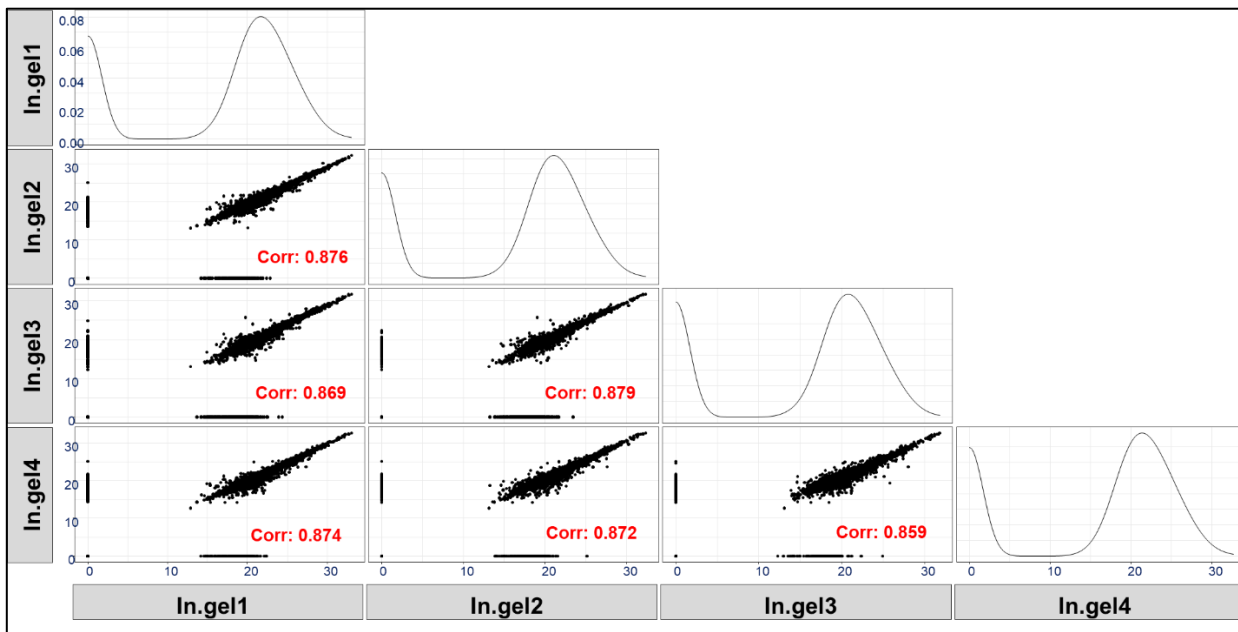
Workflow 3 (WF 3)	Same as WF 1	No connection	<u>Step 1:</u> 2µg of trypsin in 50 mM ABC (40 µl at the rate of 2 µl/min) <u>Step 2:</u> static digestion for 60 minutes with no Fluid flow	Same as WF 1	Dry the vial content in a vacuum drier, reconstitute in 0.1 % TFA	
Workflow 4 (WF 4)	Same as WF 1, 2 or 3					Invert the fluid flow in the device and collect the flowing solution (30 µl – 45 µl)
Workflow 5 (WF 5)	Same as WF 1, 2, 3 or 4	Add lysozyme in lysis buffer	Same as WF 1, 2, 3 or 4			

Annex 2 - Label-free quantification intensity for mFASP, in-gel and in-solution workflows for mix 1 and mix 3

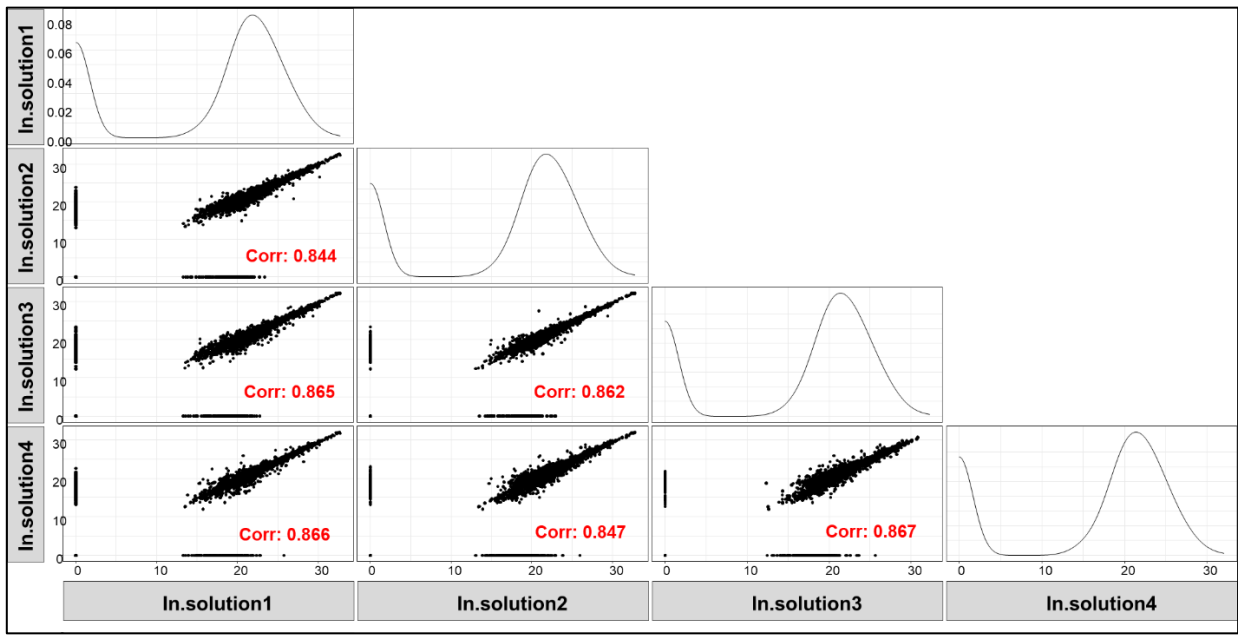
Mix 1 In-mFASP:



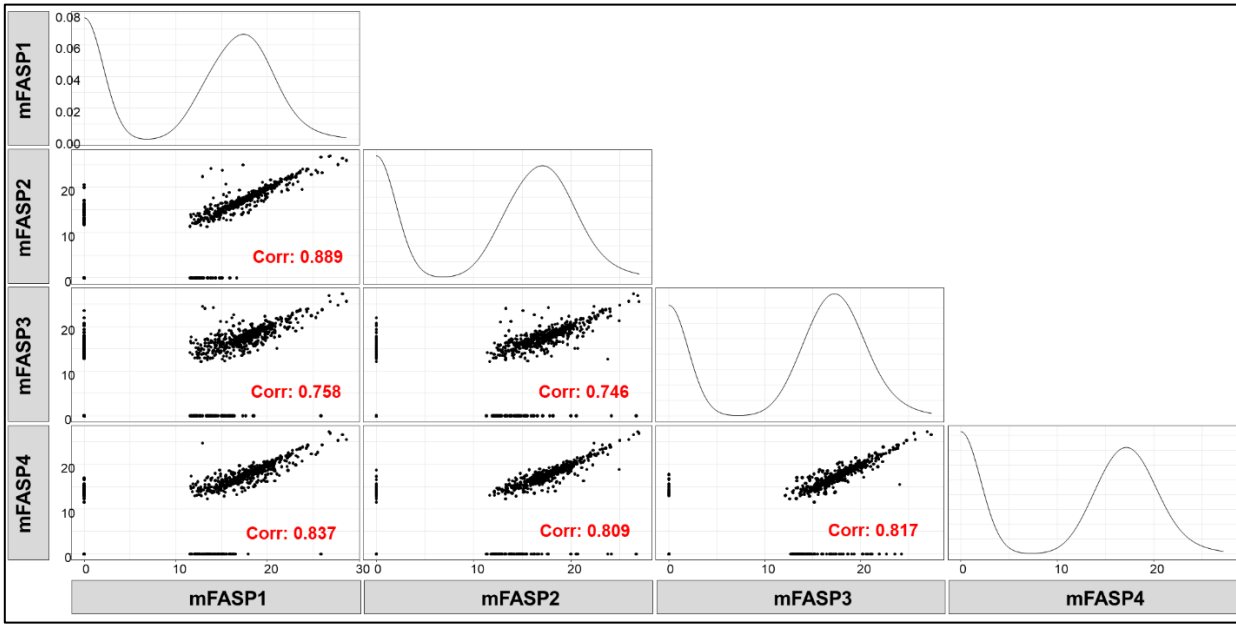
Mix 1 In-gel:



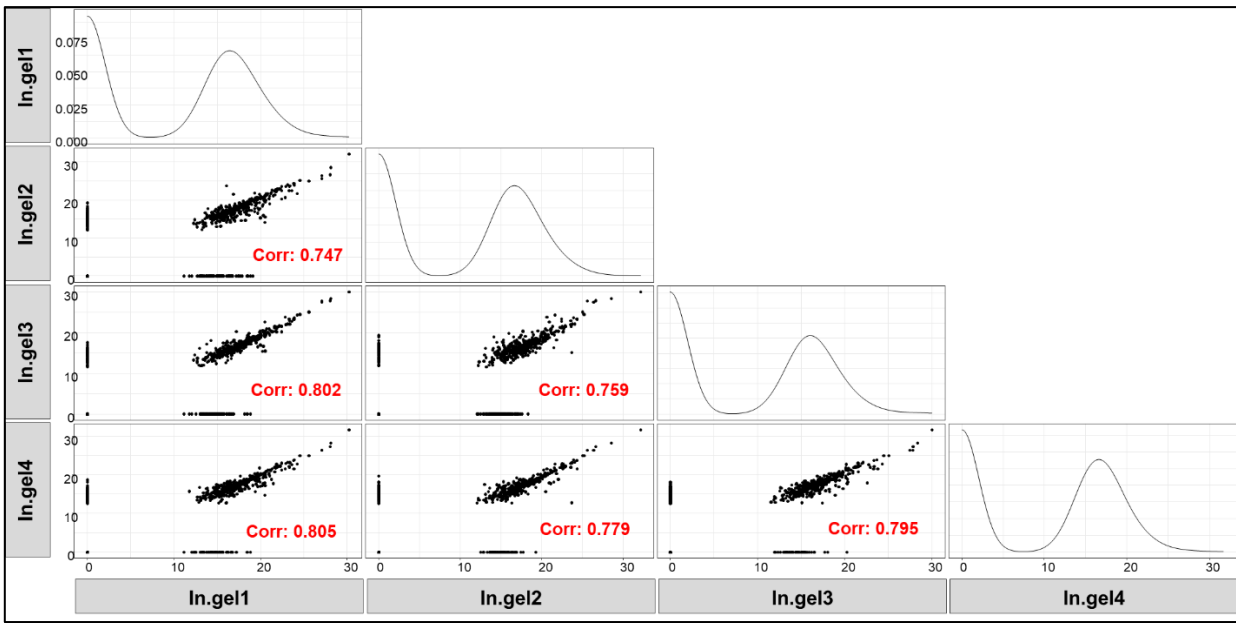
Mix 1 In-solution:



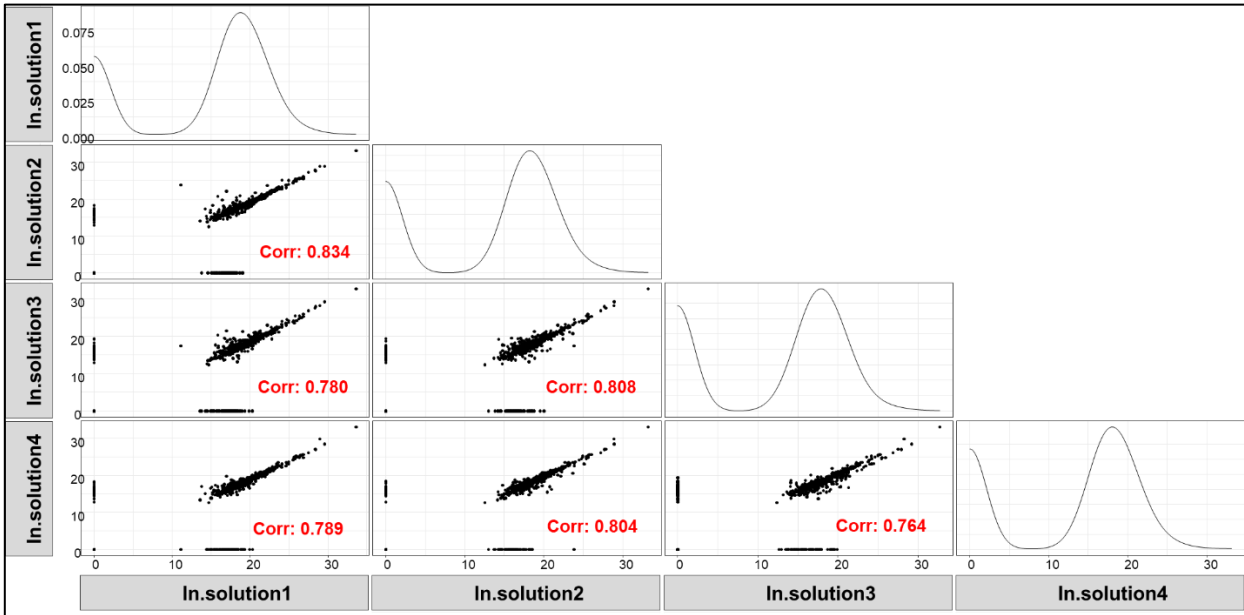
Mix 3 mFASP:



Mix 3 In-gel:



Mix 3 In-solution:



Annex 3- List of microorganisms used for building database for murine faeces microbiome data analysis

Genus	UniProt Reference
Acholeplasma	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAcholeplasma
Achromobacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAchromobacter
Agrobacterium	https://www.uniprot.org/uniprotkb?query=taxonomy_name:Agrobacterium
Akkermansia	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAkkermansia
Alistipes	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAlistipes
Anaerotruncus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAnaerotruncus
Azospirillum	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AAzospirillum
Bacillus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ABacillus
Bacteroides	https://www.uniprot.org/uniprotkb?query=taxonomy_name:Bacteroides
Bifidobacterium	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ABifidobacterium
Bilophila	https://www.uniprot.org/uniprotkb?query=taxonomy_name:Bilophila
Blautia	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ABlautia
Burkholderia	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ABurkholderia
Clostridium	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AClostridium
Coprobacillus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ACoprobacillus
Denitrovibrio	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ADenitrovibrio
Desulfovibrio	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ADesulfovibrio
Eggerthella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AEggerthella
Erysipelotrichaceae	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AErysipelotrichaceae
Escherichia	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AEscherichia
Eubacterium	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AEubacterium
Flavonifractor	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AFlavonifractor
Gloeobacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AGloeobacter
Helicobacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AHelicobacter
Lachnospiraceae	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ALachnospiraceae
Lactobacillus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ALactobacillus
Lactococcus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ALactococcus
Leuconostoc	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ALeuconostoc
Magnetospirillum	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AMagnetospirillum
Mahella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AMahella
Marvinbryantia	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AMarvinbryantia
Micavibrio	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AMicavibrio
Mus musculus	https://www.uniprot.org/uniprotkb?query=taxonomy_name:mus
Odoribacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AOdoribacter
Olsenella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AOlsenella
Oscillibacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AOscillibacter
Parabacteroides	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AParabacteroides

Parasutterella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AParasutterella
Prevotella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3APrevotella
Pseudoflavonifractor	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3APseudoflavonifractor
Pseudomonas	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3APseudomonas
Rhodospirillum	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ARhodospirillum
Roseburia	https://www.uniprot.org/uniprotkb?query=taxonomy_name:Roseburia
Ruminococcaceae	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ARuminococcaceae
Ruminococcus	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ARuminococcus
Subdoligranulum	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ASubdoligranulum
Sutterella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ASutterella
Tannerella	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ATannerella
Thermincola	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3AThermincola
Turicibacter	https://www.uniprot.org/uniprotkb?query=taxonomy_name%3ATuricibacter

**Annex 4- List of microorganisms identified in analysis of murine faeces
microbiome using ChipFilter**

Phylum	Genus	Species
Firmicutes	<i>Acetatifactor</i>	<i>Acetatifactor muris</i>
	<i>Acetivibrio</i>	<i>Acetivibrio clariflavus</i> <i>Acetivibrio thermocellus</i>
	<i>Acetoanaerobium</i>	<i>Acetoanaerobium sticklandii</i>
	<i>Agathobacter</i>	<i>Agathobacter ruminis</i>
	<i>Alkalihalobacillus</i>	<i>Alkalihalobacillus akibai</i> <i>Alkalihalobacillus clausii</i>
	<i>Alkaliphilus</i>	<i>Alkaliphilus oremlandii</i>
	<i>Amedibacillus</i>	<i>Amedibacillus dolichus</i>
	<i>Anaerobium</i>	<i>Anaerobium acetethylicum</i>
	<i>Anaerocolumna</i>	<i>Anaerocolumna aminovalerica</i> <i>Anaerocolumna cellulositytica</i> <i>Anaerocolumna chitinilytica</i> <i>Anaerocolumna jejuensis</i> <i>Anaerocolumna sedimenticola</i>
	<i>Anaerosacchariphilus</i>	<i>Anaerosacchariphilus polymeriproducens</i>
	<i>Anaerostipes</i>	<i>Anaerostipes rhamnosivorans</i> <i>Anaerostipes sp.</i>
	<i>Anaerotignum</i>	<i>Anaerotignum propionicum</i>
	<i>Anaerotruncus</i>	<i>Anaerotruncus colihominis</i> <i>Anaerotruncus sp.</i>
	<i>Bacillus</i>	<i>Bacillus aquiflavi</i> <i>Bacillus caldolyticus</i> <i>Bacillus cereus</i> <i>Bacillus glycinifermentans</i> <i>Bacillus licheniformis</i> <i>Bacillus methanolicus</i> <i>Bacillus mycoides</i> <i>Bacillus salipaludis</i> <i>Bacillus smithii</i> <i>Bacillus sp.</i> <i>Bacillus subtilis</i> <i>Bacillus thuringiensis</i> <i>Bacillus timonensis</i> <i>Bacillus toyonensis</i> <i>Bacillus wiedmannii</i> <i>Bacillus xiamenensis</i> <i>Bacillus yapensis</i>
	<i>Bilifactor</i>	<i>Bilifactor porci</i>
	<i>Blautia</i>	<i>Blautia glucerasea</i> <i>Blautia hansenii</i>

	<i>Blautia hydrogenotrophica</i>
	<i>Blautia obeum</i>
	<i>Blautia pseudococcoides</i>
	<i>Blautia sp.</i>
	<i>Blautia wexlerae</i>
<i>Butyrivibrio</i>	<i>Butyrivibrio crossotus</i>
	<i>Butyrivibrio fibrisolvens</i>
	<i>Butyrivibrio proteoclasticus</i>
	<i>Butyrivibrio sp</i>
	<i>Butyrivibrio sp.</i>
<i>Carnobacterium</i>	<i>Carnobacterium divergens</i>
<i>Catenisphaera</i>	<i>Catenisphaera adipataaccumulans</i>
<i>Cellulosilyticum</i>	<i>Cellulosilyticum lentocellum</i>
<i>Clostridium</i>	<i>Clostridium aminophilum</i>
	<i>Clostridium asparagiforme</i>
	<i>Clostridium celerecrescens</i>
	<i>Clostridium citroniae</i>
	<i>Clostridium fimetarium</i>
	<i>Clostridium hylemonae</i>
	<i>Clostridium polysaccharolyticum</i>
	<i>Clostridium populeti</i>
	<i>Clostridium spiroforme</i>
	<i>Clostridium symbiosum</i>
	<i>Clostridium thermoalcaliphilum</i>
	<i>Clostridium ultunense</i>
	<i>Clostridium acetobutylicum</i>
	<i>Clostridium acidisoli</i>
	<i>Clostridium algidicarnis</i>
	<i>Clostridium amylolyticum</i>
	<i>Clostridium baratii</i>
	<i>Clostridium beijerinckii</i>
	<i>Clostridium bolteae</i>
	<i>Clostridium bornimense</i>
	<i>Clostridium botulinum</i>
	<i>Clostridium bovifaecis</i>
	<i>Clostridium butyricum</i>
	<i>Clostridium cadaveris</i>
	<i>Clostridium carboxidivorans</i>
	<i>Clostridium cavendishii</i>
	<i>Clostridium clostridioforme</i>
	<i>Clostridium cochlearium</i>
	<i>Clostridium colicanis</i>
	<i>Clostridium collagenovorans</i>
	<i>Clostridium diolis</i>

	<i>Clostridium disporicum</i>
	<i>Clostridium estertheticum</i>
	<i>Clostridium felsineum</i>
	<i>Clostridium fermenticellae</i>
	<i>Clostridium formicaceticum</i>
	<i>Clostridium frigidicarnis</i>
	<i>Clostridium hathewayi</i>
	<i>Clostridium homopropionicum</i>
	<i>Clostridium indicum</i>
	<i>Clostridium isatidis</i>
	<i>Clostridium magnum</i>
	<i>Clostridium neonatale</i>
	<i>Clostridium nexile</i>
	<i>Clostridium novyi</i>
	<i>Clostridium perfringens</i>
	<i>Clostridium porci</i>
	<i>Clostridium puniceum</i>
	<i>Clostridium saccharobutylicum</i>
	<i>Clostridium sartagoforme</i>
	<i>Clostridium scatologenes</i>
	<i>Clostridium scindens</i>
	<i>Clostridium septicum</i>
	<i>Clostridium sp</i>
	<i>Clostridium sp.</i>
	<i>Clostridium tarantellae</i>
	<i>Clostridium tepidiprofundum</i>
	<i>Clostridium tetanomorphum</i>
	<i>Clostridium tyrobutyricum</i>
	<i>Clostridium vincentii</i>
<i>Copranaerobaculum</i>	<i>Copranaerobaculum intestinale</i>
<i>Coprobacillus</i>	<i>Coprobacillus cateniformis</i>
	<i>Coprobacillus sp.</i>
<i>Coprococcus</i>	<i>Coprococcus comes</i>
	<i>Coprococcus eutactus</i>
	<i>Coprococcus sp.</i>
<i>Cytobacillus</i>	<i>Cytobacillus firmus</i>
<i>Dielma</i>	<i>Dielma fastidiosa</i>
<i>Dorea</i>	<i>Dorea formicigenerans</i>
	<i>Dorea longicatena</i>
	<i>Dorea phocaeensis</i>
	<i>Dorea sp.</i>
<i>Dubosiella</i>	<i>Dubosiella newyorkensis</i>
<i>Eisenbergiella</i>	<i>Eisenbergiella massiliensis</i>
	<i>Eisenbergiella porci</i>

	<i>Eisenbergiella</i> sp.
	<i>Eisenbergiella tayi</i>
<i>Enterocloster</i>	<i>Enterocloster aldenensis</i>
	<i>Enterocloster asparagiformis</i>
	<i>Enterocloster bolteae</i>
	<i>Enterocloster citroniae</i>
	<i>Enterocloster clostridioformis</i>
	<i>Enterocloster lavalensis</i>
<i>Epulopiscium</i>	<i>Epulopiscium</i> sp.
	<i>Epulopiscium</i> sp.
<i>Erysipelothrix</i>	<i>Erysipelothrix rhusiopathiae</i>
	<i>Erysipelothrix</i> sp.
	<i>Erysipelothrix</i> sp.
<i>Erysipelotrichaceae</i>	<i>Erysipelotrichaceae</i> bacterium
<i>Eubacterium</i>	<i>Eubacterium cellulosolvens</i>
	<i>Eubacterium rectale</i>
	<i>Eubacterium aggregans</i>
	<i>Eubacterium barkeri</i>
	<i>Eubacterium brachy</i>
	<i>Eubacterium limosum</i>
	<i>Eubacterium nodatum</i>
	<i>Eubacterium oxidoreducens</i>
	<i>Eubacterium plexicaudatum</i>
	<i>Eubacterium ramulus</i>
	<i>Eubacterium ruminantium</i>
	<i>Eubacterium saphenum</i>
	<i>Eubacterium siraeum</i>
	<i>Eubacterium</i> sp.
	<i>Eubacterium</i> sp.
	<i>Eubacterium sulci</i>
	<i>Eubacterium ventriosum</i>
<i>Evansella</i>	<i>Evansella cellulolytica</i>
<i>Extibacter</i>	<i>Extibacter muris</i>
<i>Faecalibaculum</i>	<i>Faecalibaculum rodentium</i>
<i>Faecalicatena</i>	<i>Faecalicatena contorta</i>
	<i>Faecalicatena orotica</i>
<i>Faecalicoccus</i>	<i>Faecalicoccus acidiformans</i>
<i>Faecalimonas</i>	<i>Faecalimonas umbilicata</i>
<i>Faecalitalea</i>	<i>Faecalitalea cylindroides</i>
<i>Flavonifractor</i>	<i>Flavonifractor plautii</i>
	<i>Flavonifractor</i> sp.
<i>Floccifex</i>	<i>Floccifex porci</i>
<i>Frasingicoccus</i>	<i>Frasingicoccus caecimuris</i>
<i>Fusicatenibacter</i>	<i>Fusicatenibacter saccharivorans</i>

<i>Geobacillus</i>	<i>Geobacillus stearothermophilus</i>
<i>Gottschalkia</i>	<i>Gottschalkia purinilytica</i>
<i>Herbinix</i>	<i>Herbinix luporum</i>
<i>Hespellia</i>	<i>Hespellia stercorisuis</i>
<i>Holdemania</i>	<i>Holdemania filiformis</i>
	<i>Holdemania massiliensis</i>
<i>Hydrogenibacillus</i>	<i>Hydrogenibacillus schlegelii</i>
<i>Ileibacterium</i>	<i>Ileibacterium valens</i>
<i>Intestinibaculum</i>	<i>Intestinibaculum porci</i>
<i>Johnsonella</i>	<i>Johnsonella ignava</i>
<i>Kineothrix</i>	<i>Kineothrix alysoides</i>
<i>Lachnoanaerobaculum</i>	<i>Lachnoanaerobaculum orale</i>
	<i>Lachnoanaerobaculum saburreum</i>
	<i>Lachnoanaerobaculum sp</i>
	<i>Lachnoanaerobaculum umeaense</i>
<i>Lachnobacterium</i>	<i>Lachnobacterium bovis</i>
<i>Lachnoclostridium</i>	<i>Lachnoclostridium phytofermentans</i>
	<i>Lachnoclostridium sp</i>
	<i>Lachnoclostridium sp.</i>
<i>Lachnospira</i>	<i>Lachnospira eligens</i>
	<i>Lachnospira multipara</i>
	<i>Lachnospira pectinoschiza</i>
<i>Lachnospiraceae</i>	<i>Lachnospiraceae bacterium</i>
<i>Lachnotalea</i>	<i>Lachnotalea glycerini</i>
	<i>Lachnotalea sp.</i>
<i>Lacrimispora</i>	<i>Lacrimispora algidixylanolytica</i>
	<i>Lacrimispora celerecrescens</i>
	<i>Lacrimispora saccharolytica</i>
<i>Lacticaseibacillus</i>	<i>Lacticaseibacillus rhamnosus</i>
<i>Lactiplantibacillus</i>	<i>Lactiplantibacillus pentosus</i>
	<i>Lactiplantibacillus plantarum</i>
<i>Lactobacillus</i>	<i>Lactobacillus acetotolerans</i>
	<i>Lactobacillus amylolyticus</i>
	<i>Lactobacillus amylovorus</i>
	<i>Lactobacillus bombicola</i>
	<i>Lactobacillus casei</i>
	<i>Lactobacillus crispatus</i>
	<i>Lactobacillus delbrueckii</i>
	<i>Lactobacillus equicursoris</i>
	<i>Lactobacillus gallinarum</i>
	<i>Lactobacillus gasseri</i>
	<i>Lactobacillus hamsteri</i>
	<i>Lactobacillus helsingborgensis</i>
	<i>Lactobacillus helveticus</i>

	<i>Lactobacillus hominis</i>
	<i>Lactobacillus iners</i>
	<i>Lactobacillus intestinalis</i>
	<i>Lactobacillus jensenii</i>
	<i>Lactobacillus johnsonii</i>
	<i>Lactobacillus kefiranofaciens</i>
	<i>Lactobacillus kitasatonis</i>
	<i>Lactobacillus kullabergensis</i>
	<i>Lactobacillus panisapium</i>
	<i>Lactobacillus paragasseri</i>
	<i>Lactobacillus porci</i>
	<i>Lactobacillus rodentium</i>
	<i>Lactobacillus ruminis</i>
	<i>Lactobacillus selangorensis</i>
	<i>Lactobacillus sp</i>
	<i>Lactobacillus sp.</i>
	<i>Lactobacillus taiwanensis</i>
<i>Lactococcus</i>	<i>Lactococcus fujiensis</i>
	<i>Lactococcus raffinolactis</i>
	<i>Lactococcus sp</i>
	<i>Lactococcus termiticola</i>
<i>Latilactobacillus</i>	<i>Latilactobacillus curvatus</i>
<i>Lentilactobacillus</i>	<i>Lentilactobacillus hilgardii</i>
	<i>Lentilactobacillus kefiri</i>
<i>Leuconostoc</i>	<i>Leuconostoc citreum</i>
<i>Levilactobacillus</i>	<i>Levilactobacillus brevis</i>
<i>Ligilactobacillus</i>	<i>Ligilactobacillus ruminis</i>
	<i>Ligilactobacillus salivarius</i>
<i>Limosilactobacillus</i>	<i>Limosilactobacillus fermentum</i>
	<i>Limosilactobacillus reuteri</i>
<i>Longicatena</i>	<i>Longicatena caecimuris</i>
<i>Mahella</i>	<i>Mahella australiensis</i>
<i>Marvinbryantia</i>	<i>Marvinbryantia formatexigens</i>
<i>Massilimicrobiota</i>	<i>Massilimicrobiota sp.</i>
<i>Mediterraneibacter</i>	<i>Mediterraneibacter butyricigenes</i>
	<i>Mediterraneibacter catenae</i>
	<i>Mediterraneibacter sp.</i>
<i>Mobilisporobacter</i>	<i>Mobilisporobacter senegalensis</i>
<i>Moorella</i>	<i>Moorella thermoacetica</i>
<i>Muricomes</i>	<i>Muricomes intestini</i>
<i>Oribacterium</i>	<i>Oribacterium sp</i>
	<i>Oribacterium sp.</i>
<i>Oscillibacter</i>	<i>Oscillibacter sp</i>
	<i>Oscillibacter sp.</i>

<i>Paenibacillus</i>	<i>Paenibacillus lautus</i>
	<i>Paenibacillus polymyxa</i>
<i>Paeniclostridium</i>	<i>Paeniclostridium sordellii</i>
<i>Parasporobacterium</i>	<i>Parasporobacterium paucivorans</i>
<i>Priestia</i>	<i>Priestia megaterium</i>
<i>Pseudobutyrvibrio</i>	<i>Pseudobutyrvibrio ruminis</i>
	<i>Pseudobutyrvibrio sp.</i>
	<i>Pseudobutyrvibrio xylanivorans</i>
<i>Pseudoflavonifractor</i>	<i>Pseudoflavonifractor capillosus</i>
	<i>Pseudoflavonifractor sp.</i>
<i>Robinsoniella</i>	<i>Robinsoniella peoriensis</i>
	<i>Robinsoniella sp.</i>
<i>Roseburia</i>	<i>Roseburia hominis</i>
	<i>Roseburia intestinalis</i>
	<i>Roseburia inulinivorans</i>
	<i>Roseburia porci</i>
	<i>Roseburia sp</i>
	<i>Roseburia sp.</i>
<i>Ruminococcaceae</i>	<i>Ruminococcaceae bacterium</i>
<i>Ruminococcus</i>	<i>Ruminococcus lactaris</i>
	<i>Ruminococcus torques</i>
	<i>Ruminococcus albus</i>
	<i>Ruminococcus flavefaciens</i>
	<i>Ruminococcus gnavus</i>
	<i>Ruminococcus sp</i>
	<i>Ruminococcus sp.</i>
<i>Schaedlerella</i>	<i>Schaedlerella arabinosiphila</i>
<i>Secundilactobacillus</i>	<i>Secundilactobacillus collinoides</i>
<i>Shuttleworthia</i>	<i>Shuttleworthia sp.</i>
<i>Simiaoa</i>	<i>Simiaoa sunii</i>
<i>Solibacillus</i>	<i>Solibacillus silvestris</i>
<i>Sporofaciens</i>	<i>Sporofaciens musculi</i>
<i>Sporosarcina</i>	<i>Sporosarcina pasteurii</i>
<i>Subdoligranulum</i>	<i>Subdoligranulum sp</i>
	<i>Subdoligranulum sp.</i>
	<i>Subdoligranulum variabile</i>
<i>Suipraeoptans</i>	<i>Suipraeoptans intestinalis</i>
<i>Thermincola</i>	<i>Thermincola potens</i>
	<i>Thermoanaerobacter</i>
<i>Thermoanaerobacter</i>	<i>thermohydrosulfuricus</i>
	<i>Thermoanaerobacterium</i>
<i>Thermoanaerobacterium</i>	<i>thermosaccharolyticum</i>
<i>Thermoclostridium</i>	<i>Thermoclostridium stercorarium</i>
<i>Turicibacter</i>	<i>Turicibacter sanguinis</i>
	<i>Turicibacter sp.</i>

	<i>Tyzzarella</i>	<i>Tyzzarella sp</i>
	<i>uncultured</i>	<i>Tyzzarella sp.</i>
		<i>uncultured Anaerotruncus</i>
		<i>uncultured Blautia</i>
		<i>uncultured Clostridium</i>
		<i>uncultured Dorea</i>
		<i>uncultured Eubacterium</i>
		<i>uncultured Flavonifractor</i>
		<i>uncultured Lachnospira</i>
		<i>uncultured Oscillibacter</i>
		<i>uncultured Roseburia</i>
		<i>uncultured Ruminococcus</i>
	<i>Variimorphobacter</i>	<i>Variimorphobacter saccharofermentans</i>
	<i>Velocimicrobium</i>	<i>Velocimicrobium porci</i>
	<i>Waltera</i>	<i>Waltera intestinalis</i>
	<i>Wansuia</i>	<i>Wansuia hejianensis</i>
	<i>Weizmannia</i>	<i>Weizmannia coagulans</i>
Bacteroidetes	<i>Alistipes</i>	<i>Alistipes communis</i>
		<i>Alistipes dispar</i>
		<i>Alistipes finegoldii</i>
		<i>Alistipes indistinctus</i>
		<i>Alistipes onderdonkii</i>
		<i>Alistipes putredinis</i>
		<i>Alistipes shahii</i>
		<i>Alistipes sp</i>
	<i>Bacteroides</i>	<i>Bacteroides acidifaciens</i>
		<i>Bacteroides caccae</i>
		<i>Bacteroides caecimuris</i>
		<i>Bacteroides coprocola</i>
		<i>Bacteroides coprosuis</i>
		<i>Bacteroides fragilis</i>
		<i>Bacteroides graminisolvens</i>
		<i>Bacteroides heparinolyticus</i>
		<i>Bacteroides intestinalis</i>
		<i>Bacteroides luti</i>
		<i>Bacteroides oleiciplenus</i>
		<i>Bacteroides ovatus</i>
		<i>Bacteroides pectinophilus</i>
		<i>Bacteroides reticulotermitis</i>
		<i>Bacteroides sp</i>
		<i>Bacteroides thetaiotaomicron</i>
		<i>Bacteroides uniformis</i>
		<i>Bacteroides xylanisolvens</i>
	<i>Odoribacter</i>	<i>Odoribacter sp.</i>

	<i>Parabacteroides</i>	<i>Parabacteroides chartae</i> <i>Parabacteroides chinchillae</i> <i>Parabacteroides distasonis</i> <i>Parabacteroides gordonii</i> <i>Parabacteroides johnsonii</i> <i>Parabacteroides merdae</i> <i>Parabacteroides sp</i>
	<i>Phocaeicola</i>	<i>Phocaeicola plebeius</i> <i>Phocaeicola vulgatus</i>
	<i>Porphyromonas</i>	<i>Porphyromonas asaccharolytica</i>
	<i>Prevotella</i>	<i>Prevotella amnii</i> <i>Prevotella baroniae</i> <i>Prevotella bergensis</i> <i>Prevotella bryantii</i> <i>Prevotella buccalis</i> <i>Prevotella copri</i> <i>Prevotella corporis</i> <i>Prevotella dentalis</i> <i>Prevotella disiens</i> <i>Prevotella histicola</i> <i>Prevotella hominis</i> <i>Prevotella intermedia</i> <i>Prevotella koreensis</i> <i>Prevotella loescheii</i> <i>Prevotella maculosa</i> <i>Prevotella micans</i> <i>Prevotella pallens</i> <i>Prevotella pleuritidis</i> <i>Prevotella rara</i> <i>Prevotella ruminicola</i> <i>Prevotella shahii</i> <i>Prevotella sp</i> <i>Prevotella stercorea</i> <i>Prevotella timonensis</i>
	<i>Tannerella</i>	<i>Tannerella forsythia</i> <i>Tannerella sp.</i>
	<i>uncultured</i>	<i>uncultured Bacteroides</i>
Actinobacteria	<i>Bifidobacterium</i>	<i>Bifidobacterium adolescentis</i> <i>Bifidobacterium animalis</i> <i>Bifidobacterium asteroides</i> <i>Bifidobacterium catenulatum</i> <i>Bifidobacterium longum</i> <i>Bifidobacterium magnum</i> <i>Bifidobacterium platyrrhinorum</i>

		<i>Bifidobacterium pseudocatenulatum</i>
		<i>Bifidobacterium pseudolongum</i>
		<i>Bifidobacterium pullorum</i>
	<i>Eggerthella</i>	<i>Eggerthella guodeyinii</i>
		<i>Eggerthella lenta</i>
		<i>Eggerthella sinensis</i>
		<i>Eggerthella sp.</i>
	<i>Olsenella</i>	<i>Olsenella porci</i>
		<i>Olsenella sp.</i>
		<i>Olsenella uli</i>
Proteobacteria	<i>Achromobacter</i>	<i>Achromobacter aegrifaciens</i>
		<i>Achromobacter denitrificans</i>
		<i>Achromobacter pulmonis</i>
		<i>Achromobacter sp.</i>
	<i>Afipia</i>	<i>Afipia felis</i>
	<i>Agrobacterium</i>	<i>Agrobacterium sp</i>
	<i>Atlantibacter</i>	<i>Atlantibacter hermannii</i>
	<i>Azospirillum</i>	<i>Azospirillum sp.</i>
		<i>Azospirillum thermophilum</i>
	<i>Burkholderia</i>	<i>Burkholderia ambifaria</i>
		<i>Burkholderia anthina</i>
		<i>Burkholderia oklahomensis</i>
		<i>Burkholderia plantarii</i>
		<i>Burkholderia pseudomallei</i>
		<i>Burkholderia sp.</i>
		<i>Burkholderia territorii</i>
	<i>Desulfovibrio</i>	<i>Desulfovibrio sp</i>
	<i>Escherichia</i>	<i>Escherichia alba</i>
		<i>Escherichia albertii</i>
		<i>Escherichia coli</i>
		<i>Escherichia fergusonii</i>
		<i>Escherichia marmotae</i>
		<i>Escherichia sp</i>
	<i>Helicobacter</i>	<i>Helicobacter aurati</i>
		<i>Helicobacter winghamensis</i>
	<i>Hydrogenophaga</i>	<i>Hydrogenophaga pseudoflava</i>
	<i>Magnetospirillum</i>	<i>Magnetospirillum kuznetsovii</i>
		<i>Magnetospirillum sp.</i>
	<i>Megalodesulfovibrio</i>	<i>Megalodesulfovibrio gigas</i>
	<i>Micavibrio</i>	<i>Micavibrio sp</i>
	<i>Pseudodesulfovibrio</i>	<i>Pseudodesulfovibrio aespoensis</i>
		<i>Pseudodesulfovibrio piezophilus</i>
	<i>Pseudomonas</i>	<i>Pseudomonas aeruginosa</i>
		<i>Pseudomonas cavernae</i>

		<i>Pseudomonas coronafaciens</i>
		<i>Pseudomonas delhiensis</i>
		<i>Pseudomonas flexibilis</i>
		<i>Pseudomonas floridensis</i>
		<i>Pseudomonas fluorescens</i>
		<i>Pseudomonas marincola</i>
		<i>Pseudomonas mendocina</i>
		<i>Pseudomonas nabeulensis</i>
		<i>Pseudomonas oleovorans</i>
		<i>Pseudomonas phage</i>
		<i>Pseudomonas putida</i>
		<i>Pseudomonas sp.</i>
		<i>Pseudomonas stutzeri</i>
		<i>Pseudomonas tritici</i>
		<i>Pseudomonas zhaodongensis</i>
	<i>Ralstonia</i>	<i>Ralstonia pickettii</i>
		<i>Ralstonia solanacearum</i>
	<i>Rhizobium</i>	<i>Rhizobium giardinii</i>
		<i>Rhizobium leguminosarum</i>
		<i>Rhizobium radiobacter</i>
		<i>Rhizobium sp</i>
	<i>Rhodospirillum</i>	<i>Rhodospirillum centenum</i>
	<i>Shewanella</i>	<i>Shewanella putrefaciens</i>
	<i>Shimwellia</i>	<i>Shimwellia blattae</i>
	<i>Sutterella</i>	<i>Sutterella sp</i>
Candidatus	<i>Candidatus</i>	<i>Candidatus [Bacteroides] pericalifornicus</i>
		<i>Candidatus Galacturonibacter soehngeni</i>
Cuneatibacter	<i>Cuneatibacter</i>	<i>Cuneatibacter caecimuris</i>
Cyanobacteria	<i>Gloeobacter</i>	<i>Gloeobacter kilaueensis</i>
		<i>Gloeobacter violaceus</i>
Mycoplasmatota	<i>Acholeplasma</i>	<i>Acholeplasma sp.</i>
	<i>Haploplasma</i>	<i>Haploplasma axanthum</i>
Verrucomicrobia	<i>Akkermansia</i>	<i>Akkermansia muciniphila</i>
		<i>Akkermansia sp</i>

Annex – 5

Article

**“ChipFilter: Microfluidic Based
Comprehensive Sample Preparation
Methodology for Metaproteomics”**

Published in bioRxiv

ChipFilter: Microfluidic Based Comprehensive Sample Preparation Methodology for Metaproteomics

Ranjith Kumar Ravi Kumar¹, Massamba Mbacke Ndiaye¹, Iman
Haddad¹, Joelle Vinh¹, Yann Verdier^{1*}

¹Spectrométrie de Masse Biologique et Protéomique, LPC, UMR ESPCI CNRS 8249

10 rue Vauquelin F-75005 Paris, France

ABSTRACT

Metaproteomic approach is an attractive way to describe a microbiome at the functional level, allowing the identification and quantification of proteins across a broad dynamic range as well as detection of post-translational modifications. However, it remains relatively underutilized, mainly due to technical challenges that should be addressed, including the complexity in extracting proteins from heterogenous microbial communities.

Here, we show that a ChipFilter microfluidic device coupled to LC-MS/MS can successfully be used for identification of microbial proteins. Using cultures of *E. coli*, *B. subtilis* and *S. cerevisiae*, we have shown that it is possible to directly lyse the cells and digest the proteins in the ChipFilter to allow higher number of proteins and peptides identification than standard protocols, even at low cell density. The peptides produced are overall longer after ChipFilter digestion but show no change in their degree of hydrophobicity. Analysis of a more complex mixture of 17 species from the gut microbiome showed that the ChipFilter preparation was able to identify and estimate the amount of 16 of these species.

These results show that ChipFilter can be used for the proteomic study of microbiomes, in particular in the case of low volume or low cell density.

INTRODUCTION:

From the introduction of metaproteome and metaproteomics concepts (Rodriguez-Valera, 2004; Wilmes & Bond, 2004), studies have been done on various microbial communities. For examples, intestinal microecology, marine biology, soil biology, aerosols composition, and studies of food composition to explain food quality, safety and allergies (Yang et al., 2020). One of the major advantages of metaproteomics studies is that they provide functional information and reveal the microorganism functions and interactions at the protein level, complementary to metagenomic or metatranscriptomics data (Kleiner, 2019, Heintz-Buschart & Wilmes, 2018). Yet, metaproteomics remains relatively underutilized, mainly due to the challenges that remain in extracting proteins from heterogeneous microbial communities.

Extracting proteins from different microbial communities includes various challenges, without even mentioning sampling problems. While a universal extraction protocol providing good protein yields from a wide range of samples would be desirable, this objective does not seem achievable given the heterogeneity of matrices and microorganisms' characteristics (Keiblinger et al., 2016). Commonly, protein extraction includes a direct cellular lysis step, which is attained via chemical lysis using detergents and stabilizing agents; physical lysis (heat, pressure or snap-freezing) or enzymatic lysis. The choice of the detergent is crucial and greatly impacts on the quality of the results (Glatter et al., 2015). Proteins can be purified using different methods, such as filter-based methods FASP (Wiśniewski et al., 2009), precipitation with acids like trichloroacetic acid (TCA) (Pérez-Rodríguez et al., 2020), separation on a polyacrylamide gel using electrophoretic mobility (Granvogl et al., 2007), or solid-phase separation (Hugues et al., 2014). Each method has advantages, like protein fractionation, recovery and washing off the detergent. Then, proteins are in most cases identified by mass spectrometry (MS) using a bottom-up approach, in which proteolytic peptides are analysed to generate protein inferences. The choice of the pre-analytical strategy must consider the heterogeneity of microbial cells, having varied cellular membranes that cannot be lysed by a universal method and have species specific challenges such as a high nucleic acids contents for bacteria or wall structure for fungi. Furthermore, combining several cell lysis procedures increases the risk of losing low-abundant proteins, experimental time and handling steps. Other challenges include automation, repeatability between biological replicates and interference of detergents with the subsequent purification and analysis techniques. Therefore, for better metaproteomics sample preparation, developing a new strategy or revising existing methods is necessary.

Microfluidics offer multiple advantages in the sample preparation of microorganisms for proteomic analysis, including automation, low-volume sample handling, safety and fast processing. Microfluidic technology has been successfully applied for the separation of bacterial and viral particles from bioaerosols (Hong et al., 2015), physical cell lysis (Grigorov et al., 2021 for a review) and chemical sample processing by utilizing immobilized trypsin (Huang et al., 2006). In a previous work (Ndiaye et al., 2020), we proposed a ChipFilter Proteolysis (CFP) microfluidic device as a reactor for the miniaturization of protein sample processing and digestion steps. The CFP design is closely related to the experimental setup of filter-aided sample processing (FASP). The microchip has two reaction chambers of 0.6 μ l volume separated by a protein filtration membrane made using regenerated cellulose to concentrate or retain large polypeptides while releasing small molecules less than 10 kilodalton (kDa). Yeast protein extract and whole human cell proteome have been successfully analysed using CFP.

This study aims to assess a CFP-based workflow for sample preparation for LC-MS/MS analysis in the context of microbiology and metaproteomics. The workflow described for CFP (Ndiaye et al., 2020) was modified to introduce microbial cells directly into the device to perform all steps necessary for sample preparation starting from microbial cell lysis to proteolysis. On a mix of three microorganisms, CFP offers performance advantages compared to other methods including mFASP, in-gel and in-solution proteolysis. More proteins and peptides are identified with CFP than compared protocols, even at low cell density. The nature of the generated peptides was studied to better understand the influence of the microfluidic system in tryptic digestion. Finally, the CFP was utilized to prepare a sample mixture of 17 complex microbial species, leading to the identification of more than 10 species-specific proteins from 14 of the species.

EXPERIMENTAL PROCEDURES:

Materials-

Octyl- β -D-glucopyranoside (ODG), protease inhibitor, dithiothreitol (DTT), ammonium bicarbonate (ABC) and iodoacetamide (IAM) were purchased from Sigma-Aldrich. Trifluoroacetic acid was purchased from Thermo Fisher Scientific and acetonitrile (ACN) from Fisher Scientific.

Trypsin Gold, Mass Spectrometry Grade was from Promega. Lysozyme (50 ng/ml) was from Thermo Fisher Scientific.

Luria-Bertani (LB) agar (Merck) plates were inoculated with *Saccharomyces cerevisiae*, *Escherichia coli* and *Bacillus subtilis* and incubated at 37°C overnight. Inoculum into LB broth was made and cells were cultivated until the optical density reached 1 at 600 nm. Before harvesting, cells were counted using a glass slide under a microscope. The cells were collected by centrifugation at 2400 g for 5 minutes at room temperature. Single wash with phosphate buffer saline was performed prior to pelleting and stored at -80 °C until further use. The three cell types were considered in different cell densities for the experiments. Two different dilutions of the cells were taken as a mix of 1:1:1 cell number ratio, 10E6 of each cell type for Mix 1 and 10E2 of each cell type for Mix 2. For Mix2, 10,000-fold dilution of the mix 1 was performed with 50mM ABC.

A standard whole-cell mixture consisting of 21 representative strains from 17 species of the gut microbiota (Zymo Research, ref D6331) was divided into 10 aliquots and stored in the storage solution provided by the manufacturer at -80 °C. This standard contains 18 bacterial strains including five strains of *E. coli* (JM109, B-3008, B-2207, B-766 and B-1109), 2 fungal strains, and 1 archaeal strain in staggered abundances, theoretically ranking from 20.01% to 0.0009% considering the cell number.

ChipFilter method-

The design and fabrication methodology of the microfluidic device has been explained previously (Ndiaye et al., 2020). For the comprehensive sample preparation, cells suspended in ABC were directly loaded into the device in a total volume of 30 μ l. Cells were introduced into the ChipFilter using a piston syringe (Agilent) and syringe pump (Harvard Apparatus) maintaining a flow rate of 0.01 ml/minute. For

the sequential injection of lysis buffer 1 [(1% (w/v) ODG, protease inhibitor in 150 mM Tris-HCl pH = 8.8)], 20 mM DTT in lysis buffer 1, 50 mM IAM in lysis buffer 1 and 50 mM ABC buffer was achieved using a flow-EZ pressure module, flow controller, M-switch (Fluigent), and the software Microfluidic Automation Tool (Fluigent). The flow rate and volume were maintained in two stages at 2 μ l/minute for 45 μ l and 1 μ l/minute for 30 μ l with the upper-pressure limit at 900 mbar. Finally, proteolysis was performed by introducing 20 μ l of trypsin (final concentration of 0.1 μ g/ μ l in 50 mM ABC) at room temperature. A constant flow of 50 mM ABC was maintained for 150 minutes to ensure mixing of the proteins with trypsin. The resulting proteolytic peptides were directly transferred in the flowthrough to the sample loop of the LC. The elution volume was regulated to capacitate sample loop volume. Proteolytic peptides were finally concentrated in a trapping column (C₁₈ Pepmap, 300 μ m i.d. \times 5 mm length, Thermo Fisher Scientific).

For the standard gut microbiota, 75 μ l of the mixture corresponding to approximately 3.94×10^8 cells were thawed in ice. For cell lysis, lysis buffer 2 that has the same composition as lysis buffer 1 with supplemented lysozyme (final concentration of 0.5 mg/ml) was used. All the subsequent steps were done as described above for mixed cells.

Modified Filter-aided Sample Preparation (mFASP)-

This protocol was modified from the original FASP protocol (Wiśniewski et al., 2009) to compare the two technologies (FASP and ChipFilter) that use very similar design under identical chemical conditions. Accordingly, the samples were resuspended in lysis buffer 1 (sample/lysis buffer 1 ratio of 1:5 v/v) and incubated at 37 °C for 15 minutes. The contents are then transferred to Microcon-10 kDa Centrifugal Filter Units (Merck) and centrifuged at 15,000 g for 30 minutes to remove the flowthrough. Reduction was done with 20 mM DTT at 37 °C for 2 hours and then the tubes were centrifuged at 15,000 g for 30 minutes with the flowthrough removed. Alkylation was performed with 50 mM IAM for 2 hours in the dark at 37 °C and the tubes were centrifuged at 15,000 g for 30 minutes to remove the flowthrough. Two washes with 500 μ l of 50 mM ABC buffer were performed to remove the reagents. Digestion was performed with 2 μ g trypsin in 50 mM ABC buffer for overnight at 37 °C. Elution of peptides was done by centrifugation twice at 15,000 g for 20 minutes with 50 mM ABC to ensure maximum peptide recovery. Finally, the peptides were acidified with 0.1% (v/v) TFA. Desalting was performed with C₁₈ Zip Tips (Merck) as per manufacturer guidelines. Peptides were dried and stored in -20 °C until injection to LC.

In-gel trypsin proteolysis-

The samples were resuspended in 1X Laemmli buffer with β -Mercaptoethanol (Sigma Aldrich) and heated 5 minutes at 95 °C. 40 μ l of lysate were loaded into 12% precast gels (Bio-Rad) in Tris-Glycine- SDS running buffer at a voltage of 80V for 15 minutes. A short SDS-PAGE migration was used to restrict the proteome to a short band before separation (Hartmann et al., 2014). The gel was stained with Coomassie brilliant blue R-250 and the protein band was sliced and washed with excess double distilled water under agitation (700 rpm/minute) at room temperature. The gel was dehydrated with 100% ACN and dried using a Speedvac system. Reduction was performed with 10 mM DTT in 50 mM ABC buffer at 56°C with agitation for 30 minutes. The excess solution was removed and the gel was dehydrated again with 100% ACN. The alkylation was performed next with 55 mM IAA in a 50mM ABC buffer with agitation at 37°C for 20 minutes in the dark. The excess of solution was removed and replaced by a 50 mM ABC buffer. Dehydration was performed with 50% and 100% ACN and completely dried in the Speedvac system. 200 μ l of trypsin (12.5 ng/ μ l) was prepared with 50mM ABC buffer and rehydration of

the gel was done for 45 minutes at 4 °C. Next, the excess solution was removed and digestion was performed for overnight at 37 °C with agitation. Peptides were extracted from the gel in two steps by acidification and dehydration of the gel, using 100 µl of 1% TFA (v/v) and then 5% of formic acid (v/v) separately for 15 minutes under agitation at 37 °C. The combined solutions were then transferred to a new tube. Again, the gel was dehydrated with 100% ACN for 15 minutes under agitation at 37 °C and the solution was transferred to the elution tube. The eluate was dried in the Speedvac, resuspended in 20 µl of 1% (v/v) formic acid and sonicated for 5 minutes at 37 °C. Desalting was performed with C₁₈ zip tips as per manufacturer guidelines. Peptides were then dried and stored at -20 °C until injection to LC.

In-solution Proteolysis-

The samples were resuspended in Lysis buffer 3 (6 M Urea, 150 mM Tris, 1x Protease Inhibitors, 1% (w/v) octyl-β-D-glucopyranoside) with sample/lysis buffer 3 ratio of 1:5 (v/v). Samples were subjected to mechanical disruption with glass beads of 425-600 µm (Sigma-Aldrich) in the ratio of 1:1 (v/v). Five cycles consisting of 30 sec vortex and 60 sec in ice were repeated as bead-beating pre-treatment. The lysate was carefully removed from the beads and transferred to a new tube, where the reduction was performed with 20 mM DTT at 37 °C for 2 hours. In the same tube, alkylation was performed with 100 mM IAA at 37 °C for 2 hours in the dark. Precipitation of the proteins was performed with TCA solution (final concentration 10% v/v) for 30 minutes at 4 °C, followed by centrifugation at 17,500 g at 4 °C for 1 hour. The protein pellet was washed with 80% and then 100% (v/v) ice-cold acetone, followed by drying using speed-vac. The protein pellet was solubilized in a 50 mM ABC buffer and digestion was performed with 2 µg of trypsin for overnight at 37 °C. Acidification of the peptides was performed with 0.1% (v/v) TFA. Desalting was performed with C₁₈ zip tips as per manufacturer guidelines. Peptides were dried after desalting and stored at -20 °C until injection to LC-MS.

Liquid Chromatography and Mass Spectrometry-

Samples were analysed by nanoLC-MS/MS in data-dependent acquisition (DDA) high-energy c-trap dissociation (HCD) mode using an RSLCnano UltiMate™ 3000 System coupled to a nanoESI Q-Exactive or Q-Exactive HF mass spectrometer (Thermo Fisher Scientific).

For the ChipFilter method, the peptides recovered in the trap column were separated on a capillary reverse-phase C18 column Pepmap 75 µm i.d. × 50 cm length (ThermoFisher Scientific) at 45°C with a linear 120 minutes gradient elution from 2.5% to 60% of buffer B (water/acetonitrile/formic acid 10%: 90%: 0.1% (v/v/v)) in buffer A (water/acetonitrile/formic acid 98%: 2%: 0.1% (v/v/v)) at a fixed flow rate of 220 nL/min.

For mFASP, in-gel and in-solution methods, the dried peptides were resuspended in 7 µl of 0.1 % TFA solution (v/v) and 6 µl was used for single-shot injection. Trapping was done with C₁₈ Pepmap 300 µm i.d. × 5 mm length column (Thermo Fisher Scientific) and analysed in nanoLC MS/MS with a 120 minutes gradient as described earlier.

Mix-1 and Mix-2 samples analysis were performed with a Q-Exactive mass spectrometer operated in nanoESI at 1.7 kV. Full MS survey scan were recorded over the m/z range of 400 – 2000 with a resolution of 70,000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 100 ms. Up to 15 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 30, with a precursor isolation window at 2 m/z, resolution set at 17,500 with AGC value at 1E5 with a

maximum injection time of 120 milliseconds. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 seconds.

The standard gut microbiota was analysed with a Q-Exactive HF mass spectrometer operated in nanoESI (1.6 kV). Full MS survey scans were recorded over the m/z range of 375-1500 with a resolution of 60,000 using an automatic gain control target value (AGC) of 3E6 with a maximum injection time of 60 milliseconds. Up to 20 intense 2⁺ - 5⁺ charged ions were selected for HCD with a normalized collision energy of 28%, with precursor isolation window at 2 m/z, resolution of 15,000, AGC value of 1E5 with a maximum injection time of 60 milliseconds. The minimum MS² target value was set at 1E3 and dynamic exclusion for 20 seconds.

Data Analysis-

Spectra were processed using Proteome Discoverer (PD) v2.4 (Thermo Fisher Scientific). For mixed cells, the Mascot search engine (Matrix Science Mascot 2.2.04) was used against the all taxonomy SwissProt database (release 2022_03: 568002 sequences; 205171419 residues). For the standard gut microbiota, the SequestTM search engine was used against a dedicated sequence database (20,303 sequences containing 11,357,410 residues) specifically restricted to the species in standard gut mix from UniProtKB. A list 17 species from the standard gut mix includes - *Faecalibacterium prausnitzii*, *Roseburia hominis*, *Bifidobacterium adolescentis*, *Lactobacillus fermentum*, *Clostridioides difficile*, *Methanobrevibacter smithii*, *Enterococcus faecalis*, *Clostridium perfringens*, *Veillonella rogosae*, *Bacteroides fragilis*, *Prevotella corporis*, *Fusobacterium nucleatum*, *Akkermansia muciniphila*, *Salmonella enterica*, *Escherichia coli*, *Candida albicans* and *Saccharomyces cerevisiae* along with *homo sapiens* (1,133,353 sequences containing 506,763,197 residues). The database search was performed with the following parameters: MS and MS/MS mass tolerance 10 ppm and 0.02 Da respectively, trypsin specificity with up to 2 missed cleavages, partial Carbamidomethylation (C), Deamidation (NQ) and Oxidation (M). Proteins with at least one high confidence peptide and six amino acids were validated. Target FDR was set at 0.01. In order to perform label free quantification in PD, Minora feature detection tool was used in processing workflow and precursor ions quantifier in consensus workflow with consideration for both unique and razor peptides. The abundance value obtained was used to make the scatter plot and determine the correlation using tools in R (v 4.0.3).

The Kyte and Doolittle hydrophobicity index was calculated from a published package according to (Osorio et al., 2015).

All the experiments were performed in four replicates for mixed cells and in three replicates for standard gut mix. Statistical tests to identify significance (unpaired t test) were performed and values shown always represent mean ± standard error of mean. The mass spectrometry data have been deposited on the ProteomeXchange Consortium via the PRIDE partner repository (Perez Riverol et al., 2022) [px-submission #637484]

RESULTS:

Cell lysis can be performed in ChipFilter without pre-treatment

In the previous work (N'Diaye et al., 2020), we have shown that it is possible to perform proteolysis of eukaryotic cell samples in the ChipFilter. In order to study microorganisms, it would be advantageous to perform the cell lysis directly in the ChipFilter too. It will help to avoid contamination issues and loss of material. To determine whether it was possible to lyse microbial cells on the chip, $1E6$ cells of *E. coli*, *S. cerevisiae* and *B. subtilis* were introduced separately in to the ChipFilter. Lysis was achieved using a lysis buffer containing 1% (w/v) ODG. After CFP, peptides were eluted and identified by mass spectrometry. The results are shown in Figure 1. The number of identified proteins ranked from 824 ± 50 (*B. subtilis*), $1\,121 \pm 58$ (*E. coli*), to 1339 ± 44 (*S. cerevisiae*) suggesting that the ChipFilter cell lysis and proteolysis were efficient.

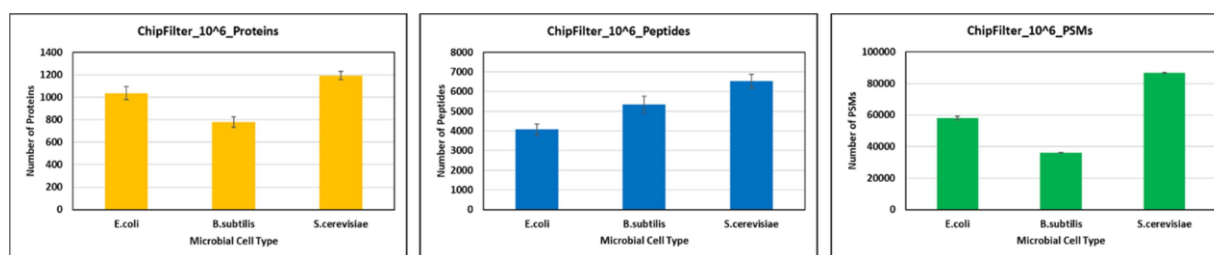


Figure 1. Whole cell lysis and bottom-up proteomic sample preparation with ChipFilter. Number of proteins (yellow), peptides (blue) and PSMs (green) identified for *Escherichia coli*, *Bacillus subtilis* and *Saccharomyces cerevisiae* using ChipFilter.

ChipFilter method identifies more proteins and peptides than other methods

In order to assess the efficiency of ChipFilter for the identification of proteins from different microorganisms, a mixture of $1E6$ *S. cerevisiae*, *E. coli* and *B. subtilis* (Mix 1) was analysed with CFP and standard proteolysis protocols (Figure 2A). The ChipFilter is a miniature of the FASP protocol, for which the fluid transfer is controlled by the laminar flow in the channels under an applied pressure, it is important to note that FASP tubes are designed for accommodating SDS based lysis buffers, but in the present study we used ODG based lysis buffers for FASP, called modified FASP (mFASP). This buffer and the lack of pre-treatment were chosen to keep it similar to the CFP. Next, an in-gel workflow is considered, as it is the most used workflow for metaproteomic sample preparation that allows the usage of SDS based lysis buffers. Finally, the in-solution method which introduces a cell lysis pre-treatment by bead beating was included to understand the influence of pre-treatment.

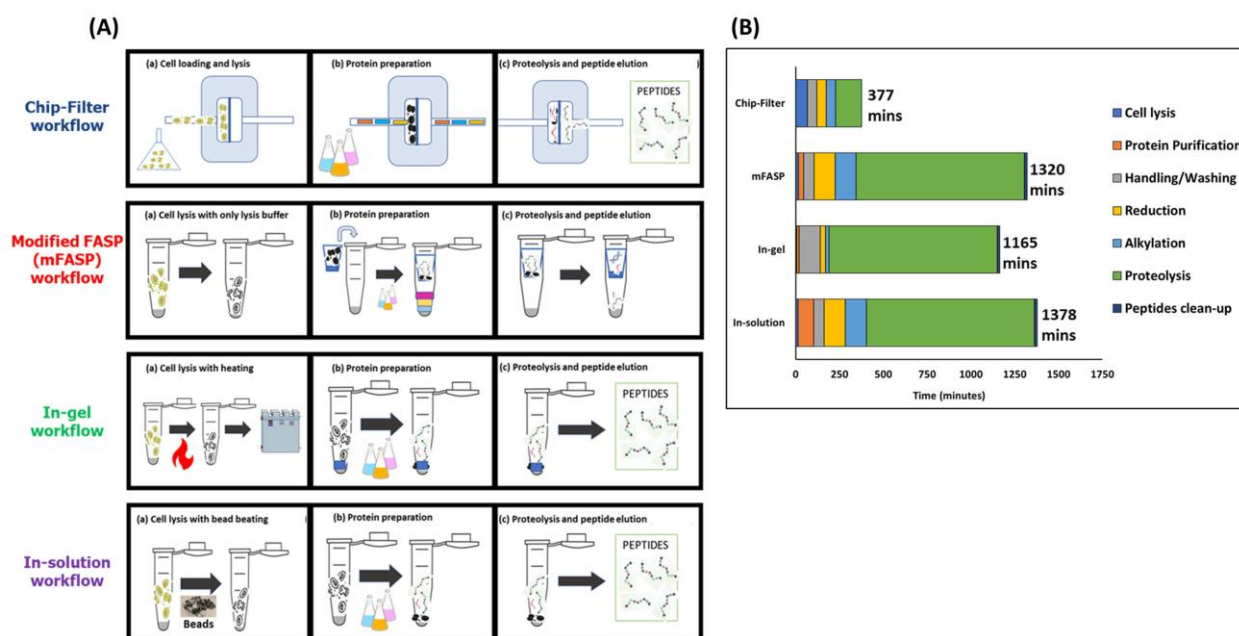


Figure 2: Comparative analysis of different workflows. (A) The different methodologies used to compare the performance of ChipFilter workflow includes modified filter assisted sample preparation (mFASP) workflow; protein separation by 12% SDS-PAGE gel and in-gel tryptic proteolysis; protein precipitation by TCA acidification and in-solution tryptic proteolysis. The steps involved in bottom-up proteomics are grouped into three key stages- cell lysis with or without pre-treatment; protein preparation including separation from other biomolecules and protein denaturation; trypsin mediated proteolysis. (B) The seven steps commonly used in bottom-up proteomics for each method along with the time taken for each step.

Figure 2 (B) highlights the time required for seven steps by each method considered. The seven steps considered include cell lysis, protein purification, handling step, reduction, alkylation, proteolysis and peptide clean-up. The digestion time for the ChipFilter is only two hours as compared to overnight for other methods, as used in most protocols. Shorter digestion time in the ChipFilter method is necessary to avoid the loss of peptides over time that exit the chamber upon digestion.

The performance of the protocols was evaluated by considering the number of target proteins and total peptides identified. For the concentration of $1E6$ cells, the ChipFilter method achieved a mean protein identification of 1999 ± 54 whereas mFASP reached 1818 ± 91 , in-gel 1504 ± 22 and in solution 1665 ± 15 (figure 3A). Similarly, at the peptide level, a significantly higher identification number is obtained for the ChipFilter method than other methods as shown in figure 3B. 9770 ± 108 peptides were identified using ChipFilter method, 8197 ± 435 with mFASP, 8218 ± 262 with in-gel and 7805 ± 178 with in-solution. Efficient catalysis in confined space and reduced loss of materials as compared to other methods can explain the higher identification for the ChipFilter method.

The distribution of the proteins and peptides among the methods show that ChipFilter protocol identifies the most distinctive proteins and peptides. At the protein level (figure 3C), 17.5 % (598 proteins) of the total proteins were identified exclusively by the ChipFilter method, whereas the other methods had less than 5.1% of specific proteins. The protein identification common between all the methods was about 38.5% (1319 proteins). At the peptide level (figure 3D), 32% (8395 peptides) of the total peptides were identified only in ChipFilter method. Interestingly, the peptide common to all the methods is only 8.8% (2309 peptides) of the total population, indicating that each method characteristically generate different peptide fractions. The ChipFilter allows the identification of 54.3% of the total peptides.

Comparison of the protein origin from three microbial cells (figure 3E) indicates that the ChipFilter identifies the highest percentage for gram-positive bacteria *B. subtilis* (12%) and gram-negative bacteria *E. coli* (37%) in comparison to other methods, whereas the highest percentage of fungi *S. cerevisiae* was recorded for in-solution (69%) method. Increased fungal proteins in the in-solution method can be reasoned due to the introduction of a pre-treatment step which was lacking in other methods. The in-gel method has poor identification for gram- positive bacteria. The usage of SDS containing Laemmli buffer can be the reason as the peptidoglycan cell wall of gram-positive bacteria was not efficiently denatured by SDS (Mahalanabis et al., 2009) . The ChipFilter method can identify all the three cell types considerably better than other methods without pre-treatment.

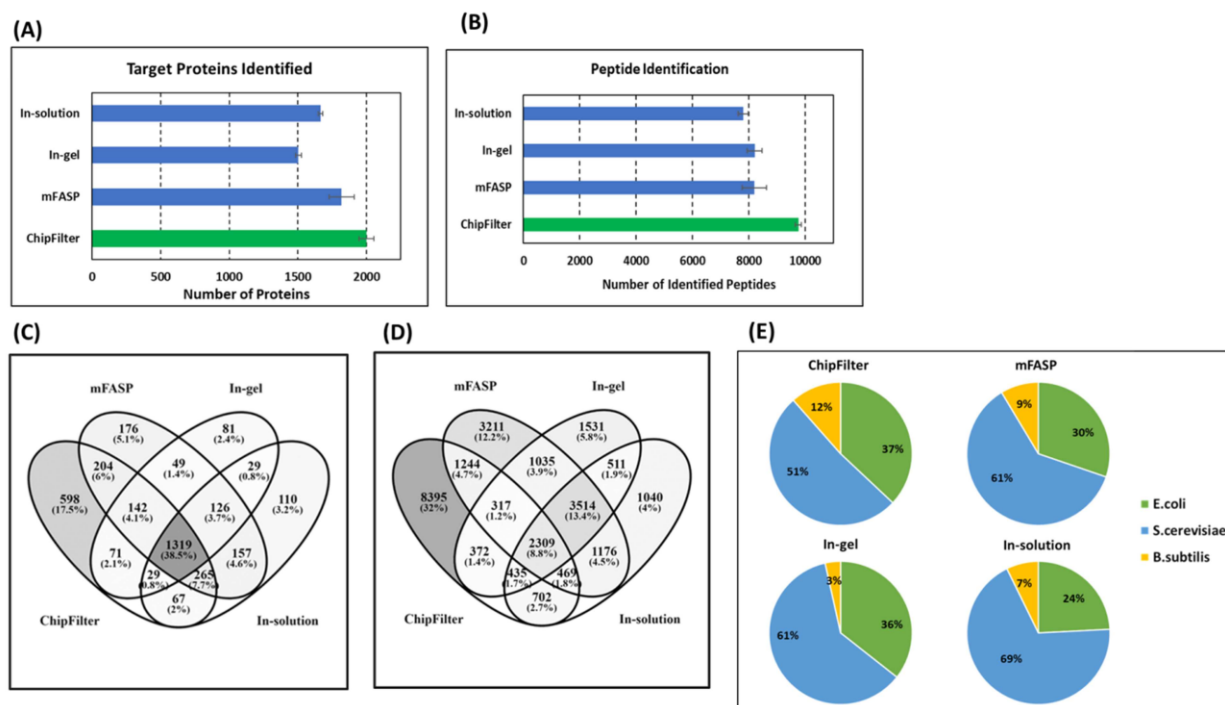


Figure 3: Performance assessment of ChipFilter workflow compared to other methods for Mix 1 (~3E6 cells/sample). Identification of (A) target proteins and (B) total peptides by different methods for four replicates. Distribution of (C) proteins and (D) peptides identified across different methods. (E) Percentage of proteins identified by each method for *E. coli*, *S. cerevisiae* and *B. subtilis*.

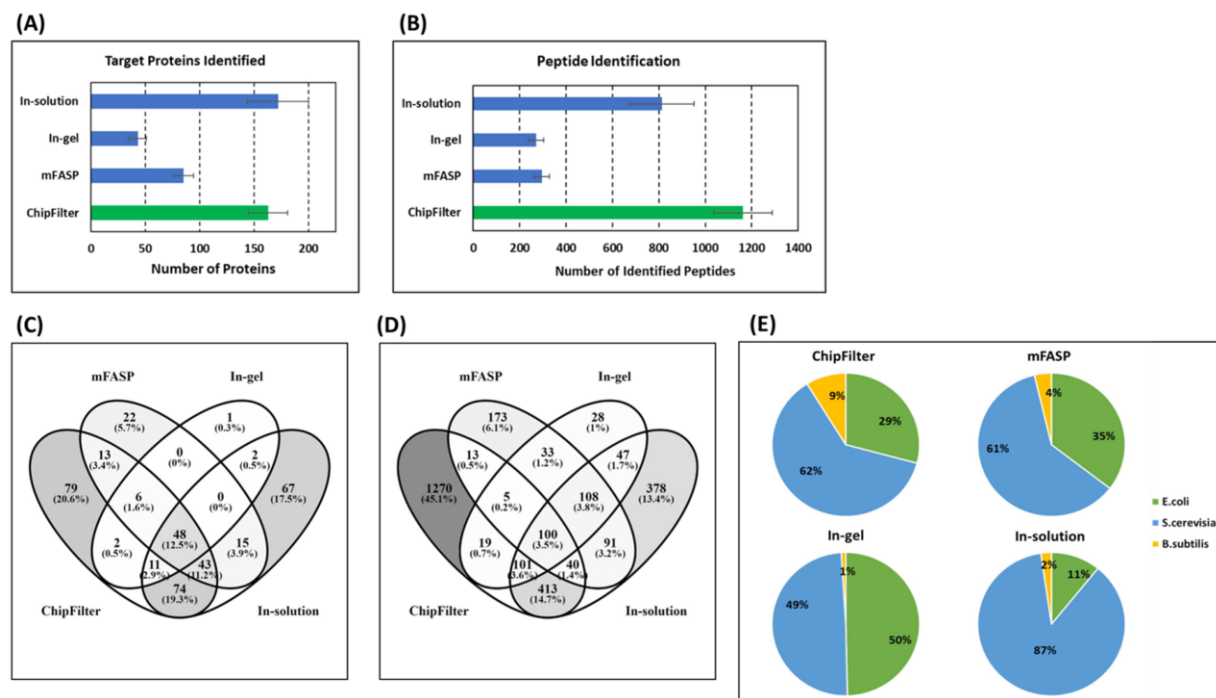


Figure 4: Performance assessment of ChipFilter workflow compared to other methods for Mix 2 (~3E2 cells/type). Identification of (A) target proteins and (B) total peptides by different methods for four replicates. Distribution of (C) proteins and (D) peptides identified across different methods. (E) Percentage of proteins identified by each method for *E. coli*, *S. cerevisiae* and *B. subtilis*.

The ChipFilter method performs efficiently with low cell number

ChipFilter was successfully used on a sample with low numbers (3E2) in mix 2 of the three microorganisms. The number of identified proteins was similar to the number obtained after in-solution proteolysis, and significantly higher than results obtained with in-gel and mFASP protocols (Figure 4A). At the peptide level, ChipFilter method allowed the identification of a greater number of peptides than other three methods tested (Figure 4B).

The sensitivity of ChipFilter method was tested for low cell number sample (Mix 2) to assess the identification and distribution of peptides and proteins like in mix 1. The mean number of target proteins and total peptides identified are 163 ± 18 and 1162 ± 126 respectively (figure 4A and 4B). The number of protein identifications is similar for ChipFilter method as compared to in-solution method (172 ± 28 proteins), but the number of peptide identification is higher than every other method.

As observed at 3E6 cell density, the coverage of identified proteins is variable depending on the protocol used. For Mix 2, the ChipFilter method identified the highest proportion of proteins (72% of the total) and the highest proportion of exclusive proteins (20.6% of the total target protein identification; Figure 4C). The same trends are obtained at the peptide level. The ChipFilter method identifies the majority of peptides (69.7%), and is the protocol that allows the identification of the largest number of exclusive peptides (45.1% of the total, Figure 4D).

The distribution of the identified proteins according to the species is more balanced with the ChipFilter protocol, in particular for *Bacillus* proteins which represent 9% of the total (9 proteins) against 2% for the other protocols (less than 4 proteins, Figure 4E)

The distribution of the proteins (figure 4C) indicates that the percentage of proteins exclusively identified by the ChipFilter method (20.6%, 79 proteins) and the in-solution method (17.5%, 67 proteins) are very close, unlike in mix 1. Furthermore, the percentage of proteins identified by both methods is about 45.9%, suggesting that the protein representation is very similar. Of the total proteins identified, only 27.9% were not identified by the ChipFilter method. For the peptide identifications shown in figure 4D, the peptides exclusively identified by the ChipFilter method is 45.1% (1270 peptides). It is roughly three times higher than the peptides exclusively identified by the in-solution method (13.4%, 378 peptides). The proportion of peptides not identified by the ChipFilter method is 30.4% (858 peptides). As observed with Mix 1, the distribution of peptides or proteins is not similar, indicating that sample preparation methods handle low cell density samples differently. Finally, the representation of protein identifications for different species with the methods is indicated in figure 4E. Like in Mix 1, the distribution of the proteins identified is represented for all three species for the ChipFilter method. In the case of the identification of *B. subtilis*, the ChipFilter identifies 9% of total proteins whereas it is less than 4% for the other methods. Furthermore, the protein identification for *S. cerevisiae* for the in-solution method is 87% of the total proteins identified which is higher than the other methods which have about 50% to 60% of this species. The ChipFilter method clearly is a universal sample preparation method for simultaneously identifying proteins in a single analysis even at low cell numbers.

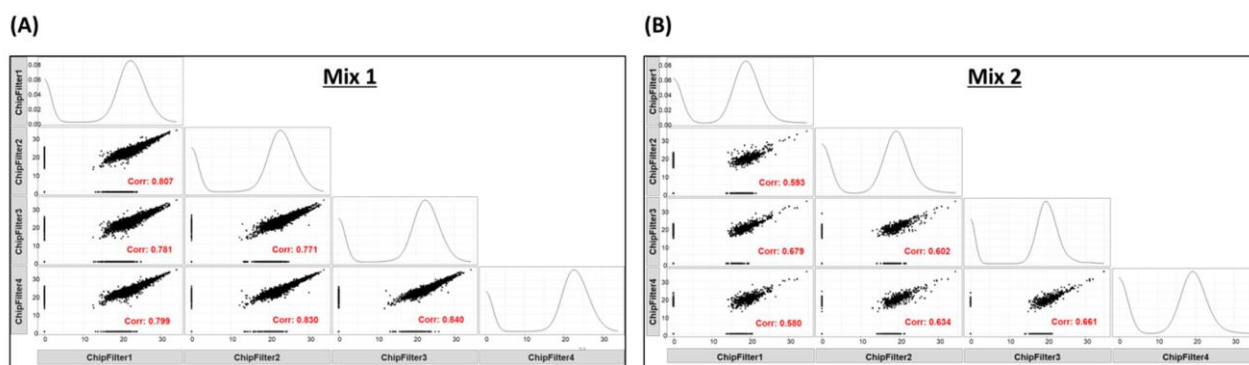


Figure 5: Association between the replicates for ChipFilter method for mix 1 and 2. The LFQ intensity is used to show the association between the different replicates (A) for mix 1

and (B) mix 2 (B). The Pearson's correlation value is indicated in red for every pair along with the scatter plot distribution.

Repeatability in ChipFilter workflow is high

To assess the repeatability of ChipFilter method, LFQ intensities among the four replicates of Mix 1 and Mix 2 cell samples have been studied. The correlation coefficient values are in the range of 0.76 to 0.83 for Mix 1 (Figure 5A), and in the range of 0.59 to 0.69 for Mix 2 (Figure 5B). The lower correlation can be due to the decreased intensities and signal-to-noise obtained for low cell numbers. These correlation coefficients are comparable to those obtained for other methods with slightly lower values for ChipFilter method.

Repeatability is often difficult to achieve in metaproteomics samples but is critical. For the current study, the variations arising during nanoLC-MS/MS are not considered and are attributed to the sample preparation methodology. Correlation is shown to be in the positive direction, indicating that changing the parameters in one replicate induces similar changes in other replicates. The correlation coefficient values for the ChipFilter method indicate a high positive correlation between the replicates for high cell numbers. Even though the peptide identifications are significantly high for the ChipFilter method in Mix 2, the correlation between the replicates make this method useable. The reduction in the correlation between the two cell densities can be due to the reduced intensities obtained at lower cell number. Comparing the correlation coefficients between the methods (data not shown) indicates that the ChipFilter method has high variability among the replicates. The in-solution and in-gel methods exhibit lower standard deviation than other methods. mFASP generates high variability like the ChipFilter method and it can be possible that the filter-based methods could contribute to these differences. In conclusion, the ChipFilter method generates replicates that have positive correlation in Mix 1 and 2 indicating that reproducibility between the different cell numbers is related to each other in the same manner.

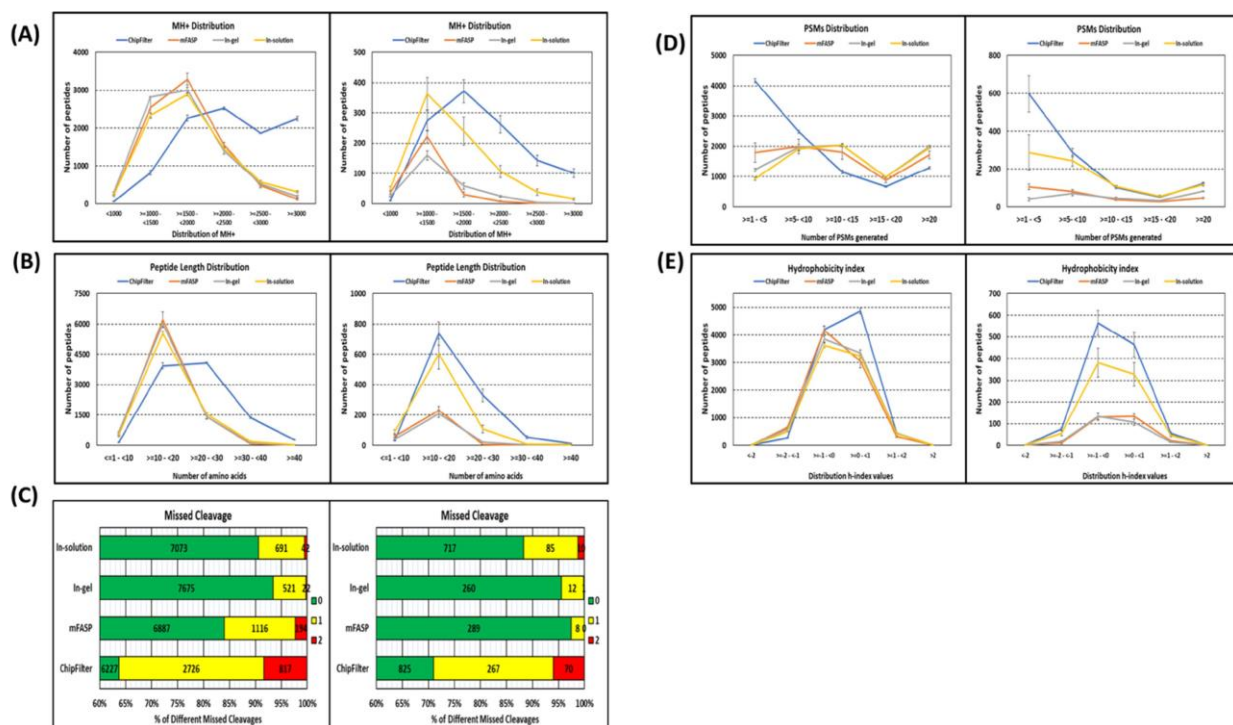


Figure 6. Physical chemical characteristics of the peptides generated by different preparation methods. (A) Positive ion mass (MH⁺) distribution for Mix 1 (left) and Mix 2 cell (right) samples. (B) Amino acid length distribution for Mix 1 (left) and Mix 2 (right) cell samples. (C) Ratio of missed cleavages generated by each method for Mix 1 (left) and Mix 2 (right) cell samples. (D) Peptide spectral matches (PSMs) distribution for Mix 1 (left) and Mix 2 (right) cell samples. (E) Kyte-Doolittle hydrophobicity index is shown in (E) for Mix 1 (left) and Mix 2 (right) samples. For (B) to (E) the mean value is plotted in the graph along with the error bars indicating the standard deviation between four replicates.

[Physio-chemical characteristics of the peptides generated by the ChipFilter method](#)

As the ChipFilter method allows the identification of a large number of specific peptides (Figure 2D and 3D), these peptides were studied to see if they have different physio-chemical characteristics from the peptides obtained by other methods, which could lead to a bias in the exploitation of the results. This comparison was done for 3E6 and 3E2 cell samples.

For the ChipFilter method with Mix 1 sample, 32% (3137 ± 149) of the peptides have a mass below 2000 Da, whereas for other methods it is approximately 73% (5879 ± 232 peptides) (Figure 6A). For Mix 2 sample, the percentage of peptides with a mass below 2000 Da was 56% in the ChipFilter. It is to be noted that the highest mass of the peptides generated was 4988 Da for a ChipFilter cut-off of 10,000 Da. These results are in line with those for amino acid size (Figure 6B). Using the ChipFilter method, 42% of Mix 1 peptides have a length of up to 20 amino

acids, instead of approximately 80% for other methods. Similarly, the proportion of Mix 2 peptides shorter than 20 amino acids was 66% with the CFP while 91% for the other methods.

To explain this result, the percentage of missed cleavage was studied for Mix 1 and Mix 2 samples (Figure 6C). The ChipFilter method has the highest percentages of 1 and 2 missed cleavages with Mix 1 (28% and 8% respectively) and Mix 2 (23% and 6% respectively). The percentage is also considerably high in the mFASP method for Mix 1 (14% and 2% respectively). In-solution method is consistent between the two cell numbers with around 10% for 1 missed cleavage and 1% for 2 missed cleavages. The in-gel method has the least. The higher proportion of missed cleavage with ChipFilter can be explained by factors such as higher density of proteins in a confined volume or shorter proteolysis duration.

Peptide spectral matches (PSMs) are useful to identify the sequence of the peptide's spectra generated in tandem MS by assigning scores to the theoretical spectra that can be generated for a specific peptide fragment. This method depends on the fragmentation of the peptides to compute the scores. Often it is considered that generation of a high number of PSMs per peptide is appropriate. On this note, the PSMs distribution was studied and is represented in figure 6D. In Mix 1, results indicated that the fraction of peptides identified with less than 10 PSMs is 6649 ± 112 (68%) for the ChipFilter method, whereas in other methods, this number dropped to 3273 ± 272 (36%). In Mix 2, the fraction of peptides under 10 PSMs is higher for the ChipFilter method (844 ± 119 peptides, 76%) than in Mix 1. Interestingly in Mix 2 except for in-gel (41%), other methods generated higher fraction of peptides under 10 PSMs (68%) The higher percentage of missed cleavages in the ChipFilter method could have contributed to the generation of peptides with a lower PSM number.

Finally, the hydrophobicity of the peptides generated was studied using the Kyte-Doolittle hydrophobicity index (figure 6E). Most of the peptides were accumulated between -1 to +1 which is indicative of a neutral mixture with slight hydrophobic or slight hydrophilic peptides. This trend was observed for all the methods including the ChipFilter method for both cell mixtures.

[ChipFilter allows efficient proteomic analysis of a standard gut microbiome](#)

In order to assess the efficiency of ChipFilter proteolysis on a more complex sample, a commercial mixture of 17 species, including bacteria (Gram-negative and Gram-positive), fungi and archaea, was used. This gut standard mixture is non-uniform in cell number and contains species commonly found in the intestinal microbiota of humans.

A mean of 2521 ± 375 proteins belonging to the 17 expected species have been identified (Figure 7) with a total of 6099 ± 1140 peptides and $18\,339 \pm 1535$ PSMs.

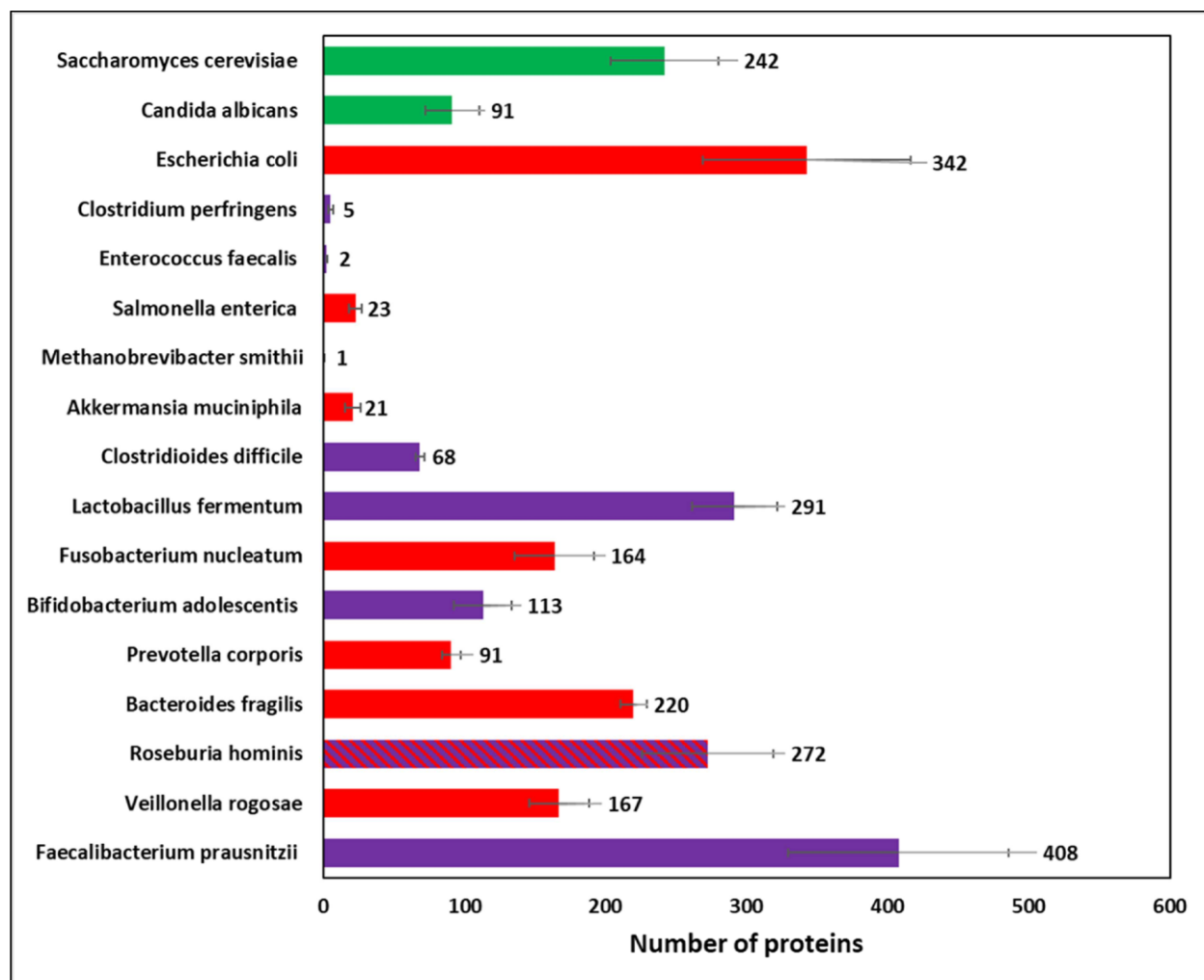


Figure 7: Number of proteins identified from each species in the standard gut mix after preparation by CFP. Green: fungi, Red: Gram-negative bacteria; Violet: Gram-positive bacteria.

R. hominis which is Gram invariable is represented in two colours. The five separate *E. coli* strains present in the mixture are grouped under a single species name.

The proteomic data were used to estimate the biomass of each species (Kleiner et al., 2017). The results are compared with the cell number, 16S and 18S and genome copy number information provided by the manufacturer (Figure 8).

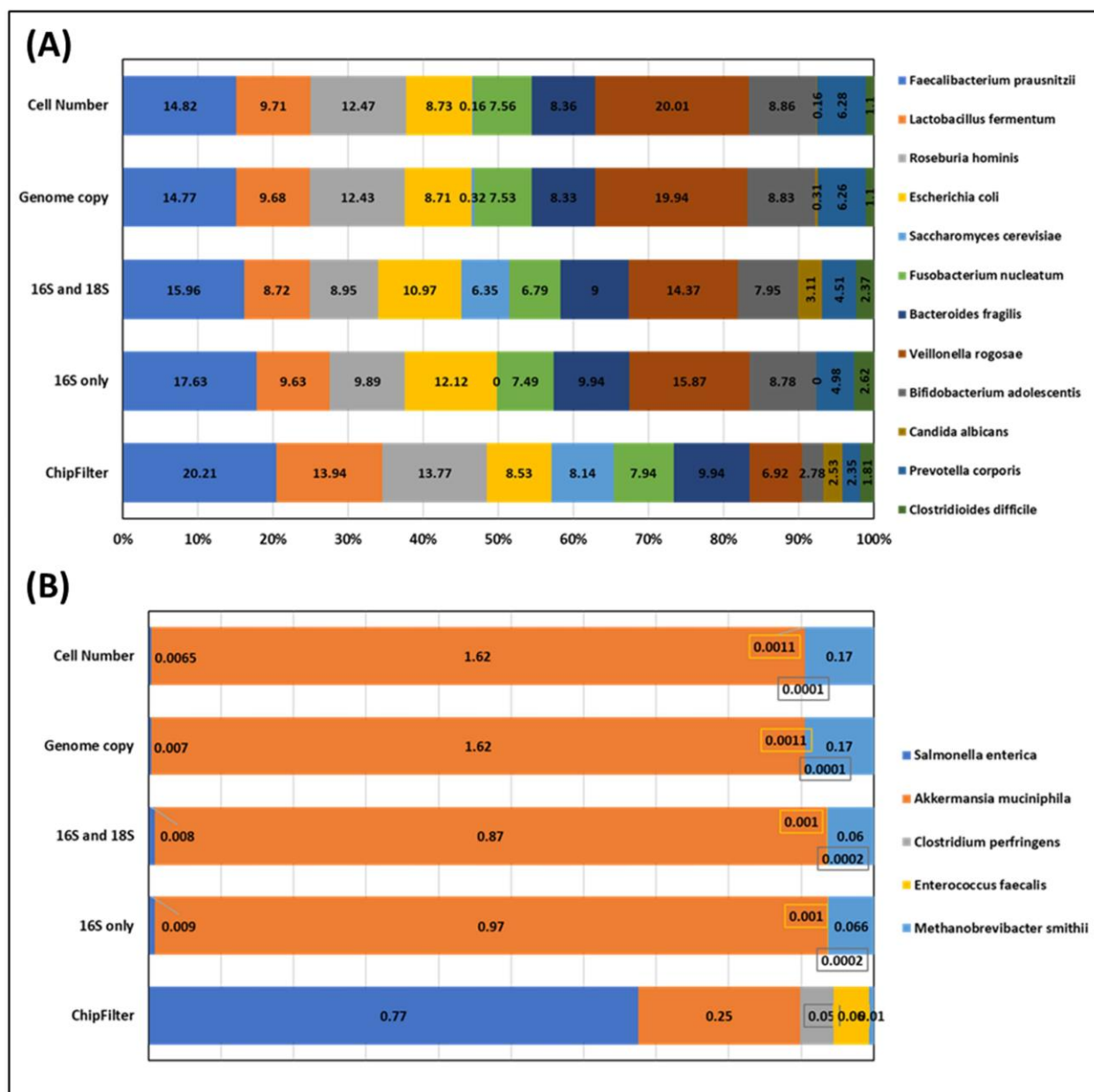


Figure 8: Biomass estimations using Next Generation Sequencing methods and ChipFilter proteomics analysis, for species present in the standard gut mix at (A) at more than 1% abundance and (B) less than 1% abundance (B). *Cell number, genome copy, 16S and 18S and 16S data were provided by the manufacturer.

DISCUSSION:

In previous work, we have shown the usefulness of the ChipFilter device for the preparation of eukaryotic cells for bottom-up proteomic analysis (Ndiaye et al., 2020). In the present study, we have adapted this system to the analysis of microbial cells, having different phenotype and cell wall compositions. We have shown that it is possible to perform cell lysis directly into the ChipFilter device, before other pre-analytical steps are performed. On a mixture of three species, we identified more proteins and peptides than standard protocols compared in the study. We have extended the use of this ChipFilter device to the analysis of a standard gut microbiome. Proteins from almost all species were identified. Moreover, these results were used to estimate the contribution of each species to the biomass, giving results in line with those determined by nucleic acid techniques.

To use the ChipFilter device for whole microbial cell sample preparation, we had to make two types of changes. The first concerns the sample loading. Our earlier work mainly focused on the usage of cellular lysates that was introduced to the ChipFilter by using the same microfluidic capillaries used for the other reagents. When using this set-up for microbial cells, we observed a cross-contamination between the samples. To overcome this problem, samples were introduced in the device using an external piston syringe and pump, and not through capillaries. Second, the lysis buffer containing ODG was adapted to be more efficient in microbial cell lysis with addition of lysozyme in case of complex mixtures comprising Gram positive bacteria. Detergents are the commonly preferred reagents for cell lysis which are often enhanced by use of physical or mechanical lysis methods to break the cell wall and membranes. Commonly used lysis buffers use anionic detergent- sodium dodecyl sulfate (SDS) in the range of 1 – 5% (m/v). SDS is highly effective in cell lysis but is generally associated with several drawbacks including requirement for urea washing to neutralize, reduction in trypsin activity and influence in mass spectrometry detection (Botelho et al., 2010). Hence, we use an alternative lysis buffer without SDS, that contains non-ionic detergent ODG. It allowed direct cell lysis on the microfluidic device without any pre-treatment. Simple washing with ammonium bicarbonate buffer was sufficient to remove the detergent from the device.

Particular attention was paid to the physio-chemical characterisation of the peptides produced after CFP. In CFP, variation in the nature of peptides generated can be caused by peptide adsorption on PDMS, continuous fluid flow and protein processing conditions provided. Peptides generated by the ChipFilter method were novel as they had higher percentage of peptides with long amino acid chains, higher mass and more missed cleavage than other methods included in the study. One of the reasons for the generation of distinct peptides can be the proteolysis condition provided inside the ChipFilter. Mechanisms of proteolysis can be broadly classified into in-gel (Shevchenko et al., 2006), in-solution (León et al., 2013) and in-solid phase (Hughes et al., 2019). While the mechanism of catalysis in the device can be argued to be a hybrid between in-solution and in-solid phase. The chamber offers the space for interaction between the proteins and the trypsin similar to in-solution, whereas the nitrocellulose

membrane and the PDMS adsorb the trypsin thereby acting as solid-phase to enable catalysis. Also, during the catalysis, a small flow rate (0.5 μ l/minute) of the ABC buffer was provided to increase the exchange of the reactants. These dual digestion mechanisms offered by the device can also be an influencing factor in the generation of distinct peptides. It is well known that shorter digestion time can generate missed cleavages. Further, shorter digestion times have been reported in past study to be advantageous (Deng et al., 2018). The ChipFilter also identifies most of the peptides identified by other methods. All this suggests that even though the number and size of peptides is greater after ChipFilter proteolysis, it does not seem to induce an identification bias compared to other methods.

The ChipFilter method was used to study standard gut microbiome consisting of 17 different species with different abundance. Our proteomic data allowed us to identify at least 10 proteins specific to 14 of these species. Proteins support the vast majority of the metabolic functions in living organisms. Knowing the protein composition of a community allows to distinguish the nature of the microorganisms that generate it and is necessary to describe functions like cell-to-cell interactions, contribution of cells in the biochemical pathways operating within a community like breakdown of food by gut communities (Wang et al., 2020) or identifying biomarkers (Heintz-Buschart et al., 2016 ; Park et al. 2021). In the present work, functional analysis of identified proteins has not been carried out due to use of standard mixture. But as a proof of concept, the contribution of each species in the overall biomass was quantified according to (Kleiner et al. 2017). In this model, the quantification data are summed based on the taxonomic assignment of inferred proteins and not based on the taxonomic assignment of peptide identifications because peptides are frequently associated with multiple proteins from different taxa. Our results are in line with those proposed by other sequencing (16S, 18S or shot-gun sequencing) results provided by the manufacturers. This aligns the proteomic finding with nucleic acid results obtained externally. It also shows the feasibility of the ChipFilter method for proteomic sample preparation.

In conclusion, the ChipFilter based sample preparation method for bottom-up proteomics by LC-MS/MS allowed in the identification of microbial proteins from single species to communities, thereby establishing the feasibility for metaproteomics. The advantage to confine low cell number cells without loss of proteins during lysis or washing stages, automation and efficient identification will allow in accelerated sample preparation with better identification. The use of ChipFilter could be extended to study microbiomes with low sample volume or cell density that will require confinement during preanalytical steps.

REFERENCES:

- Botelho D, Wall MJ, Vieira DB, Fitzsimmons S, Liu F, Doucette A. (2010) Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J Proteome Res*, 9(6): 2863–2870.
- Deng J, Julian MH, Lazar IM. (2018) Partial Enzymatic Reactions: A Missed Opportunity in Proteomics Research. *Rapid Commun Mass Spectrom*. 32(23): 2065–2073.
- Glatter T, Ahrné E, Schmidt A. (2015) Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells. *J Proteome Res*, 14(11): 4472-4485.
- Granvogl B, Plösch M, Eichacker LA. (2007) Sample preparation by in-gel digestion for mass spectrometry-based proteomics. *Analytical and Bioanalytical Chemistry*, (389): 991–1002.
- Grigorov E, Kirov B, Marinov MB, Galabov (2021). Review of Microfluidic Methods for Cellular Lysis. *Micromachines (Basel)*, 12(5): 498.
- Hartmann EM, Allain F, Gaillard JC, Pible O, Armengaud J. (2014) Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol. Biol*, 1197: 275–285.
- Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider JG, Hogan A, de Beaufort C, Wilmes P. (2016) Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol*. 2: 16180.
- Heintz-Buschart A, Wilmes P. (2018) Human Gut Microbiome: Function Matters. *Trends Microbiol*, 26(7): 563-574.
- Hong SH, Kang JS, Lee JE, Kim SS, Jung JH. (2015) Continuous aerosol size separator using inertial microfluidics and its application to airborne bacteria and viruses. *Lab Chip*, 15(8): 1889-1897.
- Huang Y, Shan W, Liu B, Liu Y, Zhang Y, Zhao Y, Lu H, Tang Y, Yang P. (2006) Zeolite nanoparticle modified microchip reactor for efficient protein digestion. *Lab Chip*, 6(4): 534-539.
- Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB,, Krijgsveld J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc*, 14(1): 68-85.
- Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol*, 10(10):757.
- Keiblinger KM., Fuchs S, Zechmeister-Boltenstern S, Riedel K. (2016) Soil and leaf litter metaproteomics- a brief guideline from sampling to understanding. *FEMS Microbiol Ecol*, 92(11): fiw180.
- Kleiner M. (2019) Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems*, 4(3):e00115-19.

- Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, Strous M. (2017) Assessing species biomass contributions in microbial communities via metaproteomics. *Nature Communications*, (8):1558
- León IR, Schwämmle V, Jensen ON, Sprenger RS. (2013) Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Mol Cell Proteomics*, 12(10): 2992–3005.
- Mahalanabis M, Al-Muayad H, Kulinski D, Altman D, Klapperichac CM. (2009) Cell lysis and DNA extraction of gram-positive and gram-negative bacteria from whole blood in a disposable microfluidic chip. *Lab Chip*, 9: 2811-2817
- Ndiaye MM, Ta HP, Chiappetta G, Vinh. (2020) On-Chip Sample Preparation Using a ChipFilter Coupled to NanoLC-MS/MS for Bottom-Up Proteomics. *J Proteome Res*, 19 (7): 2654-2663.
- Osorio D, Rondon-Villarreal P, Torres R. (2015) Peptides: A Package for Data Mining of Antimicrobial Peptides. *R Journal*, 7 (1): 4-15.
- Park J, Kim NE, Yoon H, Shin CM, Kim N, Lee DH, Park JY, Choi CH, Kim JG, Kim YK, Shin TS, Yang J, Park YS. (2021) Fecal Microbiota and Gut Microbe-Derived Extracellular Vesicles in Colorectal Cancer. *Front Oncol*, 11: 650026.
- Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA. (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 50(D1): D543–D552.
- Pérez-Rodríguez S, Ramírez OT, Trujillo-Roldán MA, Valdez-Cruz NA. (2020). Comparison of protein precipitation methods for sample preparation prior to proteomic analysis of Chinese hamster ovary cell homogenates. *Electronic Journal of Biotechnology*, (48): 86-94.
- Rodríguez-Valera F. (2004) Environmental genomics, the big picture? *FEMS Microbiology Letters*, 231(2): 153–158.
- Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* , 1: 2856–2860.
- Yang L, Fan W, Xu Y. (2020) Metaproteomics insights into traditional fermented foods and beverages. *Compr Rev Food Sci Food Saf*, 19(5): 2506-2529.
- Wang Y, Zhou Y, Xiao X, Zheng J, Zhou H. (2020) Metaproteomics: A strategy to study the taxonomy and functionality of the gut microbiota. *J Proteomics*, 219: 103737.
- Wilmes P, Bond PL. (2004) The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environmental Microbiology*, 6(9): 911–920.
- Wiśniewski JR, Zougman A, Nagaraj N, Mann M. (2009) Universal sample preparation method for proteome analysis. *Nature Methods*, (6): 359–362.

ACKNOWLEDGMENT:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 754387, the PSL University and Région Ile de France.

This work has benefited from the technical contribution of the joint service unit CNRS UAR 3750. The authors would like to thank the engineers of this unit for their advice during the development of the experiments.

References:

- Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science translational medicine*, 6, 237ra65.
- Aalismail, N. A., Ngugi, D. K., Díaz-Rúa, R., et al. (2019). Functional metagenomic analysis of dust-associated microbiomes above the Red Sea. *Sci Rep*, 9, 1–12.
- Abram, F. (2015). Systems-based approaches to unravel multispecies microbial community functioning. *Comput Struct Biotechnol J*, 13, 24–32.
- Albertsen, M., McIlroy, S.J., Stokholm-Bjerregaard, M., Karst, S.M., & Nielsen, P.H. (2016). “*Candidatus propionivibrio aalborgensis*”: a novel glycogen accumulating organism-abundant in full-scale enhanced biological phosphorus removal plants. *Front. Microbiol.*, 17.
- Alisoltani, A., Manhanzva, M. T., Potgieter, M., et al. (2020). Microbial function and genital inflammation in young South African women at high risk of HIV infection. *Microbiome*, 8, 165.
- Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., Pollard, K. S., Sakharova, E., Parks, D. H., Hugenholtz, P., Segata, N., Kyrpides, N. C., & Finn, R. D. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*, 39(1), 105–114.
- Alonso-Saez, L., Waller, A. S., Mende, D. R., Bakker, K., Farnelid, H., Yager, P. L., Lovejoy, C., Tremblay, J. E., Potvin, M., Heinrich, F., et al. (2012). Role for urea in nitrification by polar marine archaea. *Proc. Natl. Acad. Sci. USA*, 109, 17989–17994.
- Alsaeed, B., & Mansour, F. R. (2020). Distance-based paper microfluidics; principle, technical aspects and applications. *Microchemical Journal*, 155, 104664.
- Alzahrani, E., & Welham, K. (2013). Preconcentration of milk proteins using octadecylated monolithic silica microchip. *Analytica Chimica Acta*, 798, 40–47.
- Amador-García, A., Zapico, I., Borrajo, A., Malmström, J., Monteoliva, L., & Gil, C. (2021). Extending the Proteomic Characterization of *Candida albicans* Exposed to Stress and Apoptotic Inducers through Data-Independent Acquisition Mass Spectrometry. *mSystems*, 6(5), e0094621.
- Amato, P., Besaury, L., Joly, M., et al. (2019). Metatranscriptomic exploration of microbial functioning in clouds. *Sci Rep*, 9, 4383.
- Aminov, R. I. (2010). A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology*, 1, 134.

- Anderson, M. W., & Schrijver, I. (2010). Next generation DNA sequencing and the future of genomic medicine. *Genes*, 1 (1), 38–69.
- Anderson, R. C., Su, X., Bogdan, G. J., & Fenton, J. (2000). A miniature integrated device for automated multistep genetic assays. *Nucleic Acids Research*, 28(12), e60–e60.
- Archer, S., Lee, K., Caruso, T., et al. (2022). Global biogeography of atmospheric microorganisms reflects diverse recruitment and environmental filtering. *Res Sq*, PPR451906.
- Arellano, P., VanInsberghe, D., Elsherbini, J., Gore, J., & Polz, M. F. (2019). A reverse ecology approach based on a biological definition of microbial populations. *Cell*, 178, 820–34.
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., ... members), M. C. (additional. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346), 174–180.
- Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol*, 5, 782–791.
- Bailey, M. T., Dowd, S. E., Galley, J. D., Hufnagle, A. R., Allen, R. G., & Lyte, M. (2011). Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation. *Brain Behav. Immun*, 25, 397–407.
- Bajaj, J. S., Ridlon, J. M., Hylemon, P. B., Thacker, L. R., Heuman, D. M., Smith, S., Sikaroodi, M., & Gillevet, P. M. (2012). Linkage of gut microbiome with cognition in hepatic encephalopathy. *Am. J. Physiol. Gastrointest. Liver Physiol*, 302, 168–175.
- Baker, J. M., Chase, D. M., & Herbst-Kralovetz, M. M. (2018). Uterine Microbiota: Residents, Tourists, or Invaders? *Frontiers in Immunology*, 9.
- Baniasad, M., Kim, Y., Shaffer, M., Sabag-Daigle, A., Lelewi, I., Daly, R. A., Ahmer, B. M. M., Wrighton, K. C., & Wysocki, V. H. (2022). Optimization of proteomics sample preparation for identification of host and bacterial proteins in murine feces. *Analytical and bioanalytical chemistry*, 414(7), 2317–2331.
- Baohong, W., Mingfei, Y., Longxian, L., Zongxin, L., & Lanjuan, L. (2017). The Human Microbiota in Health and Disease. *Engineering*, 3, 71-82.
- Baquero, F., & Nombela, C. (2012). The microbiome as a human organ. *Clin Microbiol Infect*, Suppl 4, 2-4.
- Bashiardes, S., Zilberman-Schapira, G., & Elinav, E. (2016). Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and Biology Insights*. 10, 19–25.

- Bassis, C. M., Tang, A. L., Young, V. B., et al. (2014). The nasal cavity microbiota of healthy adults. *Microbiome*, 2, 27.
- Bastida, F., García, C., von Bergen, M., Moreno, J. L., Richnow, H. H., & Jehmlich, N. (2015). Deforestation fosters bacterial diversity and the cyanobacterial community responsible for carbon fixation processes under semiarid climate: A metaproteomics study. *Appl. Soil Ecol*, 93, 65–67.
- Bastida, F., Hernandez, T., & Garcia, C. (2014). Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods. *J. Proteom*, 101, 31–42.
- Bastida, F., Jehmlich, N., Lima, K., Morris, B. E., Richnow, H. H., Hernandez, T., von Bergen, M., & Garcia, C. (2016). The ecological and physiological responses of the microbial community from a semiarid soil to hydrocarbon contamination and its bioremediation using compost amendment. *J. Proteom*, 135, 162–169.
- Bastida, F., Nicolás, C., Moreno, J., Hernández, T., & Garcia, C. (2010). Tracing changes in the microbial community of a hydrocarbon-polluted soil by culture-dependent proteomics. *Pedosphere*, 20, 479–485.
- Becher, D., Bernhardt, J., Fuchs, S., & Riedel, K. (2013). Metaproteomics to unravel major microbial players in leaf litter and soil environments: Challenges and perspectives. *Proteomics*, 13, 2895–2909.
- Becker, L., Steglich, M., Fuchs, S., Werner, G., & Nübel, U. (2016). Comparison of six commercial kits to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. *Scientific Reports*, 6(1), 28063.
- Behzad, H., Gojobori, T., & Mineta, K. (2015). Challenges and Opportunities of Airborne Metagenomics. *Genome Biology and Evolution*, 7, 1216–1226.
- Benndorf, D., Balcke, G.U., Harms, H., & von Bergen, M. (2007). Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J*, 1, 224–234.
- Benus, R. F., van der Werf, T. S., Welling, G. W., Judd, P. A., Taylor, M. A., Harmsen, H. J., & Whelan, K. (2010). Association between *Faecalibacterium prausnitzii* and dietary fibre in colonic fermentation in healthy human subjects. *The British journal of nutrition*, 104(5), 693–700.
- Berche, P. (2012). Louis Pasteur, from crystals of life to vaccination. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 18 Suppl 5, 1–6.
- Berg, G., Rybakova, D., Fischer, F., Cernava, T., Vergès, M.-C.C., Charles, T., et al. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, 8, 103.

- Berry, S. M., Alarid, E. T., & Beebe, D. J. (2011). One-step purification of nucleic acid for gene expression analysis via Immiscible Filtration Assisted by Surface Tension (IFAST). *Lab Chip*, 11, 1747 – 1753.
- Berthier, E., Dostie, A. M., Lee, U. N., Berthier, J., & Theberge, A. B. (2019). Open Microfluidic Capillary Systems. *Analytical Chemistry*, 91(14), 8739–8750.
- Berthier, J., Brakke, K. A., & Berthier, E. (2016). *Open Microfluidics*. ISBN 9781118720936.
- Bharti, R., & Grimm, D. G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*, 22(1), 178–193.
- Bhattacharyya, S., Bershtein, S., Yan, J., Argun, T., Gilson, A. I., Trauger, S. A., & Shakhnovich, E. I. (2016). Transient protein-protein interactions perturb *E. coli* metabolome and cause gene dosage toxicity. *eLife*, 5, e20309.
- Bhattarai, Y., Muniz Pedrogo, D. A., & Kashyap, P. C. (2017). Irritable bowel syndrome: A gut microbiota-related disorder? *Am. J. Physiol. Gastrointest. Liver Physiol*, 312, 52–62.
- Biagini, F., Calvigioni, M., Lapomarda, A., Vecchione, A., Magliaro, C., De Maria, C., Montemurro, F., Celandroni, F., Mazzantini, D., Mattioli-Belmonte, M., Ghelardi, E., & Vozzi, G. (2020). A novel 3D in vitro model of the human gut microbiota. *Scientific Reports*, 10(1), 21499.
- Bian, X., Lan, Y., Wang, B., Zhang, Y. S., Liu, B., Yang, P., Zhang, W., & Qiao, L. (2016). Microfluidic Air Sampler for Highly Efficient Bacterial Aerosol Collection and Identification. *Anal Chem*, 88 (23), 11504-11512.
- Binda, C., Gibiino, G., Coluccio, C., Sbrancia, M., Dajti, E., Sinagra, E., Capurso, G., Sambri, V., Cucchetti, A., Ercolani, G., & Fabbri, C. (2022). Biliary Diseases from the Microbiome Perspective: How Microorganisms Could Change the Approach to Benign and Malignant Diseases. *Microorganisms*, 10, 312.
- Binda, C., Lopetuso, L. R., Rizzatti, G., Gibiino, G., Cennamo, V., & Gasbarrini, A. (2018). Actinobacteria: A relevant minority for the maintenance of gut homeostasis. *Digestive and liver disease: official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*, 50(5), 421–428.
- Bize, A., Cardona, L., Desmond-Le Quemener, E., Battimelli, A., Badalato, N., Bureau, C., Madigou, C., Chevret, D., Guillot, A., Monnet, V., et al. (2015). Shotgun metaproteomic profiling of biomimetic anaerobic digestion processes treating sewage sludge. *Proteomics*, 15, 3532–3543.
- Burmeister, A., & Grünberger, A. (2020). Microfluidic cultivation and analysis tools for interaction studies of microbial co-cultures. *Current Opinion in Biotechnology*, 62, 106–115.

- Burmeister, A., Hilgers, F., Langner, A., Westerwalbesloh, C., Kerkhoff, Y., Tenhaef, N., Drepper, T., Kohlheyer, D., von Lieres, E., Noack, S., & Grünberger, A. (2019). A microfluidic co-cultivation platform to investigate microbial interactions at defined microenvironments. *Lab on a Chip*, 19(1), 98–110.
- Cady, N. C., Stelick, S., Kunnnavakkam, M. V., & Batt, C. A. (2005). Real-time PCR detection of *Listeria monocytogenes* using an integrated microfluidics platform. *Sens. Actuators B*, 107, 332 – 341.
- Calabrese, F. M., Porrelli, A., Vacca, M., Comte, B., Nimptsch, K., Pinart, M., Pischon, T., Pujos-Guillot, E., & De Angelis, M. (2021). Metaproteomics Approach and Pathway Modulation in Obesity and Diabetes: A Narrative Review. *Nutrients*, 14, 47.
- Carla, M. R., Lacerda, L. H. C., & Kenneth, F. R. (2007). Metaproteomic analysis of a bacterial community response to cadmium exposure. *J. Proteome Res*, 6, 1145–1152.
- Carlo, D., Jeong, K. H., & Lee, L. P. (2003). Reagentless mechanical cell lysis by nanoscale barbs in microchannels for sample preparation. *Lab Chip*, 3, 287–291.
- Carrascosa, L. G., Sina, A. A. I., Palanisamy, R., Sepulveda, B., Otte, M. A., Rauf, S., Shiddiky, M. J. A., & Trau, M. (2014). Molecular inversion probe-based SPR biosensing for specific, label-free and real-time detection of regional DNA methylation. *Chemical Communications*, 50(27), 3585–3588.
- Casavant, B. P., Berthier, E., Theberge, A. B., Berthier, J., Montanez-Sauri, S. I., Bischel, L. L., et al. (2013). Suspended microfluidics. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (25), 10111–10116.
- Chang, H. C., & Yeo, L. (2009). *Electrokinetically Driven Microfluidics and Nanofluidics*. Cambridge University Press. ISBN: 9780521860253
- Chappell, L., Russell, A. J. C., & Voet, T. (2018). Single-Cell (Multi)omics Technologies. *Annual review of genomics and human genetics*, 19, 15–41.
- Chen, D., Mauk, M., Qiu, X., Liu, C., Kim, J., Ramprasad, S., Ongagna, S., Abrams, W. R., Malamud, D., Corstjens, P. L., & Bau, H. H. (2010). An integrated, self-contained microfluidic cassette for isolation, amplification, and detection of nucleic acids. *Biomedical microdevices*, 12(4), 705–719.
- Chen, G., Bai, R., Zhang, Y., Zhao, B., & Xiao, Y. (2022). Application of metagenomics to biological wastewater treatment. *Science of The Total Environment*, 807, 150737.
- Chen, L. L., Abbaspour, A., Mkoma, G. F., Bulik, C. M., Rück, C., & Djurfeldt, D. (2021). Gut Microbiota in Psychiatric Disorders: A Systematic Review. *Psychosomatic Medicine*, 83, 679-692.

- Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature Reviews Genetics*, 20 (6), 341–355.
- Choi, J., Hong, S. C., Kim, W., & Jung, J. H. (2017). Highly Enriched, Controllable, Continuous Aerosol Sampling Using Inertial Microfluidics and Its Application to Real-Time Detection of Airborne Bacteria. *ACS sensors*, 2 (4), 513-521.
- Chokkalingam, V., Tel, J., Wimmers, F., Liu, X., Semenov, S., Thiele, J., Figdor, C. G., & Huck, W. T. (2013). Probing cellular heterogeneity in cytokine-secreting immune cells using droplet-based microfluidics. *Lab on a chip*, 13(24), 4740–4744.
- Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep*, 5, 16350.
- Chu, M., & Zhang, X. (2022). Bacterial Atlas of Mouse Gut Microbiota. *Cellular Microbiology*, 2022, 5968814.
- Coudron, L., McDonnell, M. B., Munro, I., McCluskey, D. K., Johnston, I. D., Tan, C. K. L., & Tracey, M. C. (2019). Fully integrated digital microfluidics platform for automated immunoassay; A versatile tool for rapid, specific detection of a wide range of pathogens. *Biosensors and Bioelectronics*, 128, 52-60.
- Coughlan, L., Cotter, P., Hill, C., & Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in Microbiology*, 6.
- Cui, L., Morris, A., & Ghedin, E. (2013). The human mycobiome in health and disease. *Genome Medicine*, 5 (7), 63.
- Cvetkovic, A., Menon, A. L., Thorgersen, M. P., Scott, J. W., Poole, F. L., 2nd, Jenney, F. E., Jr., Lancaster, W. A., Praissman, J. L., Shanmukh, S., Vaccaro, B. J., et al. (2010). Microbial metalloproteomes are largely uncharacterized. *Nature*, 466, 779–782.
- Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R. H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P. H., & Wagner, M. (2015). Complete nitrification by *Nitrospira* bacteria. *Nature*, 528(7583), 504–509.
- Damit, B. (2016). Droplet-based microfluidics detector for bioaerosol detection. *Aerosol Science and Technology*, 51 (4), 488-500.
- de Araujo, G.G., Rodrigues, F., Gonçalves, F. L. T., et al. (2019). Survival and ice nucleation activity of *Pseudomonas syringae* strains exposed to simulated high-altitude atmospheric conditions. *Sci Rep*, 9, 7768.

- De Weirdt, R., Possemiers, S., Vermeulen, G., Moerdijk-Poortvliet, T. C. W., Boschker, H. T. S., Verstraete, W., & Van de Wiele, T. (2010). Human faecal microbiota display variable patterns of glycerol metabolism. *FEMS Microbiology Ecology*, 74(3), 601–611.
- Dean, B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U.S.A.*, 99(8), 5261–5266.
- Denef, V. J., VerBerkmoes, N. C., Shah, M. B., Abraham, P., Lefsrud, M., Hettich, R. L., & Banfield, J. F. (2009). Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ. Microbiol*, 11, 313–325.
- Derrien, M., Vaughan, E. E., Plugge, C. M., & de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International journal of systematic and evolutionary microbiology*, 54(Pt 5), 1469–1476.
- Dertinger, S., Chiu, D., Jeon, N., & Whitesides, G. (2001). Generation of gradients having complex shapes using microfluidic networks. *Anal. Chem*, 73, 1240–46.
- Dinan, T. G., Stanton, C., & Cryan, J. F. (2013) Psychobiotics: a novel class of psychotropic. *Biol Psychiatry*, 74, 720–726.
- Dominianni, C., Wu, J., Hayes, R. B., et al. (2014). Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol*, 14, 103.
- Dommergue, A., Amato, P., Tignat-Perrier, R., Magand, O., Thollot, A., Joly, M., Bouvier, L., Sellegri, K., Vogel, T., Sonke, J. E., Jaffrezo, J.-L., Andrade, M., Moreno, I., Labuschagne, C., Martin, L., Zhang, Q., & Larose, C. (2019). Methods to Investigate the Global Atmospheric Microbiome. *Frontiers in Microbiology*, 10.
- Drabovich, A. P., Pavlou, M. P., Batruch, I., & Diamandis, E. P. (2013). Chapter 2 - Proteomic and mass spectrometry technologies for biomarker discovery (H. J. Issaq & T. D. B. T.-P. and M. A. to B. D. (Second E. Veenstra (Eds.); pp. 17–37). Academic Press.
- Drewes, J. L., Chen, J., Markham, N. O., Knippel, R. J., Domingue, J. C., Tam, A. J., Chan, J. L., Kim, L., McMann, M., Stevens, C., Dejea, C. M., Tomkovich, S., Michel, J., White, J. R., Mohammad, F., Campodónico, V. L., Heiser, C. N., Wu, X., Wu, S., ... Sears, C. L. (2022). Human Colon Cancer–Derived *Clostridioides difficile* Strains Drive Colonic Tumorigenesis in Mice. *Cancer Discovery*, 12(8), 1873–1885.
- Duarte, G. R. M., Price, C. W., Augustine, B. H., Carrilho, E., & Landers, J. P. (2011). Dynamic Solid Phase DNA Extraction and PCR Amplification in Polyester-Toner Based Microchip. *Anal. Chem.*, 83, 5182 – 5189.

- Dykstra, A. B., Flick, T. G., Lee, B., Blue, L. E., & Angell, N. (2021). Chip-Based Capillary Zone Electrophoresis Mass Spectrometry for Rapid Resolution and Quantitation of Critical Quality Attributes in Protein Biotherapeutics. *Journal of the American Society for Mass Spectrometry*, 32(8), 1952–1963.
- Dziomba, S., Araya-Farias, M., Smadja, C., Taverna, M., Carbonnier, B., & Tran, N. T. (2017). Solid supports for extraction and preconcentration of proteins and peptides in microfluidic devices: A review. *Analytica Chimica Acta*, 955, 1–26.
- Easley, C. J., Karlinsey, J. M., Bienvenue, J. M., Legendre, L. A., Roper, M. G., Feldman, S. H., Hughes, M. A., Hewlett, E. L., Merkel, T. J., Ferrance, J. P., et al. (2006). A fully integrated microfluidic genetic analysis system with sample-in-answer-out capability. *Proc. Natl. Acad. Sci. USA*, 103, 19272 – 19277
- Easterly, C.W., Sajulga, R., Mehta, S., Johnson, J., Kumar, P., Hubler, S., et al. (2019). MetaQuantome: An integrated, quantitative metaproteomics approach reveals connections between taxonomy and protein function in complex microbiomes. *Mol Cell Proteomics*, 1, 82–91.
- Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., & Lee, L.-H. (2015). Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance. *Frontiers in Life Science*, 8(3), 284–293.
- Ericson, C., Holm, J., Ericson, T., & Hjertén, S. (2000). Electroosmosis- and Pressure-Driven Chromatography in Chips Using Continuous Beds. *Analytical Chemistry*, 72(1), 81–87.
- Fan, Y., & Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol*, 19, 55–71.
- Fjelstrup, S., Andersen, M. B., Thomsen, J., Wang, J., Stougaard, M., Pedersen, F. S., Ho, Y. P., Hede, M. S., & Knudsen, B. R. (2017). The Effects of Dithiothreitol on DNA. *Sensors (Basel, Switzerland)*, 17(6), 1201.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (New York, N.Y.)*, 269(5223), 496–512.
- Flemming, H.-C. (1993). Biofilms and Environmental Protection. *Water Science and Technology*, 27(7–8), 1–10.
- Flint, H. J., & Duncan, S. H. (2014). *Bacteroides and Prevotella* (C. A. Batt & M. L. B. T.-E. of F. M. (Second E. Tortorello (Eds.); pp. 203–208). Academic Press.
- Florens, L., Carozza, M. J., Swanson, S. K., Fournier, M., Coleman, M. K., Workman, J. L., et al. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods*, 40, 303–311.

- Floyd, D. E., Tuste, C., Jacques, I., Bruce, J. P., Anne C, R, T., Wen-Han, Y., Abirami, L., & William, G. W. (2010). The Human Oral Microbiome. *Journal of Bacteriology*, 192, 5002–5017.
- Fountain, A. W., Klemm, R., Becker, H., Hlawatsch, N., Julich, S., Miethe, P., Moche, C., Schattschneider, S., Tomaso, H., & Gärtner, C. (2014). A microfluidic platform with integrated arrays for immunologic assays for biological pathogen detection. *Proceedings of the SPIE*, 9073, 7.
- Fradique, R., Azevedo, A. M., Chu, V., Conde, J. P., & Aires-Barros, M. R. (2020). Microfluidic platform for rapid screening of bacterial cell lysis. *Journal of Chromatography A*, 1610, 460539.
- France, M., Alizadeh, M., Brown, S., et al. (2022). Towards a deeper understanding of the vaginal microbiota. *Nat Microbiol*, 7, 367–378.
- Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. USA*, 104, 13780–13785.
- Frey, K. G., Herrera-Galeano, J. E., Redden, C. L., Luu, T. V., Servetas, S. L., Mateczun, A. J., Mokashi, V. P., & Bishop-Lilly, K. A. (2014). Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC genomics*, 15, 96.
- Fricke, W.F., & Rasko, D.A. (2014). Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet*, 15, 49–55.
- Fu, X., Zhang, Y., Xu, Q., Sun, X., & Meng, F. (2021). Recent Advances on Sorting Methods of High-Throughput Droplet-Based Microfluidics in Enzyme Directed Evolution. *Frontiers in Chemistry*, 9.
- Fu, Y., Li, C., Lu, S., Zhou, W., Tang, F., Xie, X. S., & Huang, Y. (2015). Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proceedings of the National Academy of Sciences*, 112(38), 11923–11928.
- Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K. W., & Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME J*, 8, 1301–1313.
- Golden, D. M., Jay, J. M., & Loessner, M. J. (2005). *Modern food microbiology*. Berlin: Springer. page 179. ISBN 0-387-23180-3
- Gole, J., Gore, A., Richards, A., et al. (2013). Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat Biotechnol*, 31, 1126–1132.

- Goltsman, D. S., Deneff, V. J., Singer, S. W., VerBerkmoes, N. C., Lefsrud, M., Mueller, R. S., Dick, G. J., Sun, C. L., Wheeler, K. E., Zemla, A., et al. (2009). Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing “*Leptospirillum rubrum*” (group II) and “*Leptospirillum ferrodiazotrophum*” (group III) bacteria in acid mine drainage biofilms. *Appl. Environ. Microbiol.*, 75, 4599–4615.
- Goodman, S. R. B. T.-G. M. C. B. (Fourth E. (Ed.). (2021). Chapter 15 - The Microbiome. Academic Press, 393–404.
- Gow, N. A. R., & Yadav, B. (2017). Microbe Profile: *Candida albicans*: a shape-changing, opportunistic pathogenic fungus of humans. *Microbiology (Reading, England)*, 163(8), 1145–1147.
- Grassl, N., Kulak, N. A., Pichler, G., et al. (2016). Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med*, 8(1), 44.
- Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., NISC Comparative Sequencing Program, Bouffard, G. G., Blakesley, R. W., Murray, P. R., Green, E. D., Turner, M. L., & Segre, J. A. (2009). Topographical and temporal diversity of the human skin microbiome. *Science (New York, N.Y.)*, 324(5931), 1190–1192.
- Grigorov, E., Kirov, B., Marinov, M. B., & Galabov, V. (2021). Review of Microfluidic Methods for Cellular Lysis. *Micromachines*, 12(5), 498.
- Gryp, T., Glorieux, G., Joossens, M., & Vanechoutte, M. (2020). Comparison of five assays for DNA extraction from bacterial cells in human faecal samples. *Journal of applied microbiology*, 129(2), 378–388.
- Guckenberger, D. J., de Groot, T. E., Wan, A. M., Beebe, D. J., & Young, E. W. (2015). Micromilling: a method for ultra-rapid prototyping of plastic microfluidic devices. *Lab on a Chip*, 15 (11), 2364–2378.
- Gutierrez, D. B., Gant-Branum, R. L., Romer, C. E., Farrow, M. A., Allen, J. L., Dahal, N., Nei, Y.-W., Codreanu, S. G., Jordan, A. T., Palmer, L. D., Sherrod, S. D., McLean, J. A., Skaar, E. P., Norris, J. L., & Caprioli, R. M. (2018). An Integrated, High-Throughput Strategy for Multiomic Systems Level Analysis. *Journal of Proteome Research*, 17(10), 3396–3408.
- Hagan, K. A., Reedy, C. R., Uchimoto, M. L., Basu, D., Engel, D. A., & Landers, J. P. (2011). An integrated, valveless system for microfluidic purification and reverse transcription-PCR amplification of RNA for detection of infectious agents. *Lab Chip*, 11, 957 – 961.
- HaileMariam, M., Egeuz, R. V., Singh, H., Bekele, S., Ameni, G., Pieper, R., & Yu, Y. (2018). S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *Journal of Proteome Research*, 17(9), 2917–2924.
- Hamann, E., Gruber-Vodicka, H., Kleiner, M., et al. (2016). Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature*, 534, 254–258.

- Han, J. Y., Wiederoder, M., & DeVoe, D. L. (2019). Isolation of intact bacteria from the blood by selective cell lysis in a microfluidic porous silica monolith. *Microsyst. Nanoeng*, 5, 30.
- Hartmann, E. M., Allain, F., Gaillard, J. C., Pible, O., & Armengaud, J. (2014). Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods in molecular biology* (Clifton, N.J.), 1197, 275–285.
- Hasan, N., & Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 7:e7502.
- Hashmi, I., Bindschedler, S., & Junier, P. (2020). Chapter 18 - Firmicutes (N. Amaresan, M. Senthil Kumar, K. Annapurna, K. Kumar, & A. B. T.-B. M. in A.-E. Sankaranarayanan (Eds.);). Academic Press. 363–396.
- Hayoun, K., Gouveia, D., Grenga, L., Pible, O., Armengaud, J., & Alpha-Bazin, B. (2019). Evaluation of Sample Preparation Methods for Fast Proteotyping of Microorganisms by Tandem Mass Spectrometry. *Frontiers in Microbiology*, 10.
- He, S., Kunin, V., Haynes, M., Martin, H. G., Ivanova, N., Rohwer, F., Hugenholtz, P., & McMahon, K. D. (2010). Metatranscriptomic array analysis of 'Candidatus Accumulibacter phosphatis'-enriched enhanced biological phosphorus removal sludge. *Environmental Microbiology*, 12, 1205–1217.
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107, 1–8.
- Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., et al. (2019). A robust and universal metaproteomics workflow for research studies and routine diagnostics within 24 h using phenol extraction, FASP digest, and the MetaProteomeAnalyzer. *Front Microbiol*, 101883.
- Ho, H., & Bunyavanich, S. (2018). Role of the microbiome in food allergy. *Curr Allergy Asthma Rep*, 18, 27.
- Honaas, L. A., Altman, N. S., & Krzywinski, M. (2016). Study design for sequencing studies. *Methods Mol Biol*, 1418, 39–66.
- Hong, S. C., Kang, J. S., Lee, J. E., Kim, S. S., & Jung, J. H. (2015). Continuous aerosol size separator using inertial microfluidics and its application to airborne bacteria and viruses. *Lab on a chip*, 15 (8), 1889- 97.
- Hopfner, F., Künstner, A., Müller, S. H., Künzel, S., Zeuner, K. E., Margraf, N. G., Deuschl, G., Baines, J. F., & Kuhlenbäumer, G. (2017). Gut microbiota in Parkinson disease in a northern German cohort. *Brain Res.*, 1667, 41–45.

- Howe, A. C., Jansson, J. K., Malfatti, S. A., et al. (2014). Tackling soil diversity with the assembly of large, complex metagenomes. *Proc Natl Acad Sci U S A*, 111, 4904–4909.
- Hu, C., & Jia, W. (2021). Multi-omics profiling: the way towards precision medicine in metabolic diseases. *Journal of molecular cell biology*, 13(8), 576–593.
- Hu, Y., An, Q., Sheu, K., Trejo, B., Fan, S., & Guo, Y. (2018). Single Cell Multi-Omics Technology: Methodology and Application. *Frontiers in Cell and Developmental Biology*, 6.
- Huang, S., Li, C., Lin, B., & Qin, J. (2010). Microvalve and micropump controlled shuttle flow microfluidic device for rapid DNA hybridization. *Lab Chip*, 10, 2925–2931.
- Huang, X., Xing, X., Ng, C. N., & Yobas, L. (2019). Single-Cell Point Constrictions for Reagent-Free High-Throughput Mechanical Lysis and Intact Nuclei Isolation. *Micromachines*, 10, 488.
- Hughes, C. S., Moggridge, S., Müller, T., et al. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc*, 14, 68–85.
- Hughes, C., Ma, B., & Lajoie, G. A. (2010). De novo sequencing methods in proteomics. *Methods in molecular biology (Clifton, N.J.)*, 604, 105–121.
- Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., Harden, J., Turetsky, M. R., McGuire, A. D., Shah, M. B., VerBerkmoes, N. C., Lee, L. H., Mavrommatis, K., & Jansson, J. K. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*, 521(7551), 208–212.
- Hutchison, C. A., & Venter, J. C. (2006). Single-cell genomics. *Nat Biotechnol*, 24(6), 657–658.
- Inami, H., Tsuge, K., Matsuzawa, M., Sasaki, Y., Togashi, S., Komano, A., & Seto, Y. (2009). Semi-automated bacterial spore detection system with micro-fluidic chips for aerosol collection, spore treatment and ICAN DNA detection. *Biosensors & bioelectronics*, 24 (11), 3299-305.
- Islam, M. S., Aryasomayajula, T. A., & Selvaganapathy, P. R. (2017). A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*, 8, 83.
- Jansson, J. K., & Baker, E. S. (2016). A multi-omic future for microbiome studies. *Nature Microbiology*, 1(5), 16049.
- Jehmlich, N., Vogt, C., Lünsmann, V., Richnow, H. H., & von Bergen, M. (2016). Protein-SIP in environmental studies. *Current Opinion in Biotechnology*, 41, 26–33.
- Jeon, J. S., Chung, S., Kamm, R. D., & Charest, J. L. (2011). Hot embossing for fabrication of a microfluidic 3D cell culture platform. *Biomedical Microdevices*, 13 (2), 325–333.

- Jeon, N., Dertinger, S., Chiu, D., Choi, I., Stroock, A., & Whitesides, G. (2000). Generation of solution and surface gradients using microfluidic systems. *Langmuir*, 16, 8311–8316.
- Jeong, K., Kim, S., & Bandeira, N. (2012). False discovery rates in spectral identification. *BMC bioinformatics*, 13 Suppl 16(Suppl 16), S2.
- Jevtić, Ž., Stoll, B., Pfeiffer, F., Sharma, K., Urlaub, H., Marchfelder, A., & Lenz, C. (2019). The Response of *Haloferax volcanii* to Salt and Temperature Stress: A Proteome Study by Label-Free Mass Spectrometry. *PROTEOMICS*, 19(20), 1800491.
- Jiang, X., Jing, W., Sun, X., Liu, Q., Yang, C., Liu, S., Qin, K., & Sui, G. (2016). High-Throughput Microfluidic Device for LAMP Analysis of Airborne Bacteria. *ACS sensors*, 1(7), 958-962.
- Jiang, X., Jing, W., Zheng, L., Liu, S., Wu, W., & Sui, G. (2014). A continuous-flow high-throughput microfluidic device for airborne bacteria PCR detection. *Lab on a chip*, 14(4), 671-676.
- Jing, W., Jiang, X., Zhao, W., Liu, S., Cheng, X., & Sui, G. (2014). Microfluidic platform for direct capture and analysis of airborne *Mycobacterium tuberculosis*. *Anal Chem*, 86(12), 5815-21.
- Jing, W., Zhao, W., Liu, S., Li, L., Tsai, C. T., Fan, X., Wu, W., Li, J., Yang, X., & Sui, G. (2013). Microfluidic device for efficient airborne bacteria captures and enrichment. *Anal Chem*, 85(10), 5255-62.
- Kachel, S., Zhou, Y., Scharfer, P., Vrančić, C., Petrich, W., & Schabel, W. (2014). Evaporation from open microchannel grooves. *Lab on a Chip*, 14(4), 771–778.
- Kaigala, G. V., Lovchik, R. D., & Delamarche, E. (2012). Microfluidics in the "open space" for performing localized chemistry on biological interfaces. *Angewandte Chemie*, 51(45), 11224–11240.
- Kang, J. S., Lee, K. S., Kim, S. S., Bae, G. N., & Jung, J. H. (2014). Real-time detection of an airborne microorganism using inertial impaction and mini-fluorescent microscopy. *Lab on a chip*, 14(1), 244- 251.
- Kashkary, L., Kemp, C., Shaw, K. J., Greenway, G. M., & Haswell, S. J. (2012). Improved DNA extraction efficiency from low level cell numbers using a silica monolith based micro fluidic device. *Anal. Chim. Acta*, 750, 127 – 131.
- Kim, H., Bartsch, M. S., Renzi, R. F., He, J., Van de Vreugde, J. L., Claudnic, M. R., & Patel, K. D. (2011). Automated Digital Microfluidic Sample Preparation for Next-Generation DNA Sequencing. *JALA: Journal of the Association for Laboratory Automation*, 16(6), 405–414.

- Kim, J., Mauk, M., Chen, D., Qiu, X., Kim, J., Gale, B., & Bau, H. H. (2010 (A)). An integrated, self-contained microfluidic cassette for isolation, amplification, and detection of nucleic acids. *Analyst*, 135, 2408 – 2414.
- Kim, S. J., Song, Y.-A., & Han, J. (2010.B). Nanofluidic concentration devices for biomolecules utilizing ion concentration polarization: theory, fabrication, and applications. *Chemical Society Reviews*, 39(3), 912–922.
- Kim, S.-C., & Carlson, K. (2007). Temporal and Spatial Trends in the Occurrence of Human and Veterinary Antibiotics in Aqueous and River Sediment Matrices. *Environmental Science & Technology*, 41, 50–57.
- Kim, Y., Kang, J., Park, S. J., Yoon, E. S., & Park, J. K. (2007). Microfluidic biomechanical device for compressive cell stimulation and lysis. *Sens. Actuators B*, 128, 108–116.
- Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *MSystems*, 4(3), e00115-19.
- Kleiner, M., Dong, X., Hinzke, T., Wippler, J., Thorson, E., Mayer, B., & Strous, M. (2018). Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proceedings of the National Academy of Sciences*, 115(24), 5576–5584.
- Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., & Strous, M. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nature communications*, 8(1), 1558.
- Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., ... Dubilier, N. (2012). Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences*, 109(19), E1173–E1182.
- Könönen, E., & Gursoy, U. K. (2022). Oral *Prevotella* Species and Their Connection to Events of Clinical Relevance in Gastrointestinal and Respiratory Tracts. *Frontiers in microbiology*, 12, 798763.
- Kremer, C., Witte, C., Neale, S.L., Reboud, J., Barrett, M.P., & Cooper, J.M. (2015). Shape-Dependent Optoelectronic Cell Lysis. *Angew. Chem. Int. Ed.*, 53, 842–846.
- Krych, L., Hansen, C. H.F., Hansen, A. K., van den Berg, F. W. J., & Nielsen, D. S. (2013) Quantitatively Different, yet Qualitatively Alike: A Meta-Analysis of the Mouse Core Gut Microbiome with a View towards the Human Gut Microbiome. *PLoS ONE*, 8(5): e62578
- Kumamoto, C. A. (2002). *Candida* biofilms. *Current opinion in microbiology*, 5(6), 608–611.

- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews*, 72, 557–578.
- Kwon, H. J., Fronczek, C. F., Angus, S. V., Nicolini, A. M., & Yoon, J. Y. (2014). Rapid and Sensitive Detection of H1N1/2009 Virus from Aerosol Samples with a Microfluidic Immunosensor. *Journal of laboratory automation*, 19 (3), 322-331.
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1), 105–132.
- Lade, R. K., Jochem, K. S., Macosko, C. W., & Francis, L. F. (2018). Capillary Coatings: Flow and Drying Dynamics in Open Microchannels. *Langmuir*, 34 (26), 7624–7639.
- Lane, N. (2015). The unseen world: reflections on Leeuwenhoek (1677). Concerning little animals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1666), 20140344.
- Larsen, N., Vogensen, F. K., van den Berg, F. W., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., Al-Soud, W. A., Sorensen, S. J., Hansen, L. H., & Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE*, 5, 9085.
- Laukens, D., Brinkman, B. M., Raes, J., et al. (2016). Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. *FEMS Microbiol Rev*, 40, 117–132.
- Lauro, F. M., DeMaere, M. Z., Yau, S., Brown, M. V., Ng, C., Wilkins, D., Raftery, M. J., Gibson, J. A., Andrews-Pfannkoch, C., Lewis, M., et al. (2011). An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J*, 5, 879–895.
- Lederberg, J., & Mccray, A. T. (2001). `Ome Sweet `Omics--A genealogical treasury of words. *The Scientist*, 15(7), 8–8.
- Lee, C. Y., Lee, G. B., Lin, L., Huang, F. C., & Liao, C. J. (2005). Integrated microfluidic systems for cell lysis, mixing/pumping, and DNA amplification. *J. Micromech. Microeng*, 15, 1215–1223.
- Lee, H., Kim, J., Lim, H., Cho, E., Huh, N., Ko, C., Park, J., Choi, J., & Lee, S. (2010). Electrochemical cell lysis device for DNA extraction. *Lab Chip*, 10, 626–633.
- Lee, J. Y., Jin, H. S., Kim, K. S., Baek, J. H., Kim, B. S., & Lee, D. W. (2022). Nutrient-specific proteomic analysis of the mucin degrading bacterium *Akkermansia muciniphila*. *Proteomics*, 22(3), e2100125.
- Legendre, L. A., Bienvenue, J. M., Roper, M. G., Ferrance, J. P., & Landers, J. P. (2006). A Simple, Valveless Microfluidic Sample Preparation Device for Extraction and Amplification of DNA from Nanoliter-Volume Samples. *Anal. Chem*, 78, 1444 – 1451.

- Leipert, J., & Tholey, A. (2019). Miniaturized sample preparation on a digital microfluidics device for sensitive bottom-up microproteomics of mammalian cells using magnetic beads and mass spectrometry-compatible surfactants. *Lab on a Chip*, 19(20), 3490–3498.
- Leung, K., Zahn, H., Leaver, T., Konwar, K. M., Hanson, N. W., Pagé, A. P., Lo, C.-C., Chain, P. S., Hallam, S. J., & Hansen, C. L. (2012). A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proceedings of the National Academy of Sciences*, 109(20), 7665–7670.
- Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. USA*, 102, 11070–11075.
- Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122), 1022–1023.
- Li, C., Boban, M., & Tuteja, A. (2017.A). Open-channel, water-in-oil emulsification in paper-based microfluidic devices. *Lab on a Chip*, 17 (8), 1436–1441.
- Li, J., & Assmann, S. M. (2000). Mass spectrometry. An essential tool in proteome analysis. *Plant physiology*, 123(3), 807–809.
- Li, J., Smith, L. S., & Zhu, H.-J. (2021). Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. *Drug Discovery Today: Technologies*, 39, 49–56.
- Li, Q., Han, Y., Dy, A. B. C., & Hagerman, R. J. (2017.B). The gut microbiota and autism spectrum disorders. *Front. Cell. Neurosci*, 11, 120.
- Li, R., Lv, X., Zhang, X., Saeed, O., & Deng, Y. (2016). Microfluidics for cell-cell interactions: A review. *Frontiers of Chemical Science and Engineering*, 10(1), 90–98.
- Li, X., Zhang, X., Liu, Q., Zhao, W., Liu, S., & Sui, G. (2018). Microfluidic System for Rapid Detection of Airborne Pathogenic Fungal Spores. *ACS sensors*, 3 (10), 2095-2103.
- Li, Z., Wang, Y., Yao, Q., Justice, N. B., et al. (2014). Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat. Commun*, 5, 1–11.
- Lin, H.-C., Liu, Y.-J., & Yao, D.-J. (2010). Core—Shell Droplets for Parallel DNA Ligation of an Ultra-Micro Volume Using an EWOD Microfluidic System. *JALA: Journal of the Association for Laboratory Automation*, 15(3), 210–215.
- Liu, D., Liang, G., Zhang, Q., & Chen, B. (2013). Detection of *Mycobacterium tuberculosis* Using a Capillary-Array Microsystem with Integrated DNA Extraction, Loop-Mediated Isothermal Amplification, and Fluorescence Detection. *Anal. Chem.*, 85, 4698 – 4704.

- Liu, F., Lai, S., Reinmuth-Selzle, K., Scheel, J. F., Fröhlich-Nowoisky, J., Després, V. R., Hoffmann, T., Pöschl, U., & Kampf, C. J. (2016). Metaproteomic analysis of atmospheric aerosol samples. *Analytical and bioanalytical chemistry*, 408(23), 6337–6348.
- Liu, Q., Zhang, X., Li, X., Liu, S., & Sui, G. (2018.A). A semi-quantitative method for point-of-care assessments of specific pathogenic bioaerosols using a portable microfluidics-based device. *Journal of aerosol science*, 115, 173-180.
- Liu, Q., Zhang, X., Yao, Y., Jing, W., Liu, S., & Sui, G. (2018.B). A novel microfluidic module for rapid detection of airborne and waterborne pathogens. *Sensors and Actuators B: Chemical*, 258, 1138-1145.
- Liu, Z., Banaei, N., & Ren, K. (2017). Microfluidics for combating antimicrobial resistance. *Trends Biotechnol*, 35 (12), 1129– 1139.
- Lo, I., Deneff, V. J., Verberkmoes, N. C., Shah, M. B., Goltsman, D., DiBartolo, G., Tyson, G. W., Allen, E. E., Ram, R. J., Detter, J. C., et al. (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature*, 446, 537–541.
- Locascio, L., Perso, C., & Lee, C. (1999). Measurement of electroosmotic flow in plastic imprinted microfluid devices and the effect of protein adsorption on flow rate. *J. Chromatogr. A*, 857, 275–84.
- Lombard, N., Prestat, E., van Elsas, J. D., & Simonet, P. (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiology Ecology*, 78, 31–49.
- Long, S., Yang, Y., Shen, C., Wang, Y., Deng, A., Qin, Q., & Qiao, L. (2020). Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ biofilms and microbiomes*, 6(1), 14.
- Lu, H., Mutafooulos, K., Heyman, J. A., Spink, P., Shen, L., Wang, C., Franke, T., & Weitz, D. A. (2019). Rapid additive-free bacteria lysis using traveling surface acoustic waves in microfluidic channels. *Lab Chip*, 19, 4064.
- Luckey, T. D. (1972). Introduction to intestinal microecology. *The American Journal of Clinical Nutrition*, 25(12), 1292-1294.
- Luis, G. (2021). Microbiome-based diagnostics and biomarkers: Changing the paradigm in microbiology and medicine. *Microbiome times*.
- Lundberg, R., Bahl, M. I., Licht, T. R., et al. (2017). Microbiota composition of simultaneously colonized mice housed under either a gnotobiotic isolator or individually ventilated cage regime. *Sci Rep*, 7, 42245.
- Ma, Z., Zheng, Y., Cheng, Y., Xie, S., Ye, X., & Yao, M. (2016). Development of an integrated microfluidic electrostatic sampler for bioaerosol. *Journal of aerosol science*, 95, 84-94.

- Macaulay, I. C., Ponting, C. P., & Voet, T. (2017). Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends in Genetics*, 33(2), 155–168.
- Magne, F., Gotteland, M., Gauthier, L., Zazueta, A., Pesoa, S., Navarrete, P., & Balamurugan, R. (2020). The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in Obese Patients?. *Nutrients*, 12(5), 1474.
- Mahalanabis, M., Al-Muayad, H., Kulinski, M. D., Altman, D., & Klapperich, C. M. (2009). Cell lysis and DNA extraction of gram-positive and gram-negative bacteria from whole blood in a disposable microfluidic chip. *Lab on a Chip*, 9(19), 2811–2817.
- Maljkovic Berry, I., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., Morton, L., & Jarman, R. G. (2019). Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. *The Journal of Infectious Diseases*, 221 (Suppl 3), 292–307.
- Manz, A., Effenhauser, C., Burggraf, N., Harrison, D., Seiler, K., & Fluri, K. (1994). Electroosmotic pumping and electrophoretic separations for miniaturized chemical analysis systems. *J. Micromech. Microeng*, 4, 257–65.
- Marasco, G., Di Biase, A. R., Schiumerini, R., Eusebi, L. H., Iughetti, L., Ravaioli, F., Scaioli, E., Colecchia, A., & Festi, D. (2016). Gut microbiota and celiac disease. *Dig. Dis. Sci.*, 61, 1461–1472.
- Marbouty, M., & Koszul, R. (2015). Metagenome Analysis Exploiting High-Throughput Chromosome Conformation Capture (3C) Data. *Trends in genetics: TIG*, 31(12), 673–682.
- Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, 3, 31.
- Marcy, Y., Ishoey, T., Lasken, R. S., Stockwell, T. B., Walenz, B. P., Halpern, A. L., et al. (2007) Nanoliter Reactors Improve Multiple Displacement Amplification of Genomes from Single Cells. *PLoS Genet*, 3(9), e155.
- Marzorati, M., Van Den Abbeele, P., Grootaert, C., De Weirtdt, R., Carcavilla, A.M., Vermeiren, J., et al. (2013). Models of the human microbiome in vitro. In: Marchesi JR, editor. *The human microbiota and microbiome*. Oxfordshire, UK: CABI publishing;
- Mernier, G., Piacentini, N., Braschler, T., Demierre, N., & Renauda, P. (2010). Continuous-flow electrical lysis device with integrated control by dielectrophoretic cell sorting. *Lab Chip*, 10, 2077–2082.
- Meyer, F., Fritz, A., Deng, Z. L., Koslicki, D., Lesker, T. R., Gurevich, A., Robertson, G., Alser, M., Antipov, D., Beghini, F., Bertrand, D., Brito, J. J., Brown, C. T., Buchmann, J., Buluç, A., Chen, B., Chikhi, R., Clausen, P. T. L. C., Cristian, A., Dabrowski, P. W., ... McHardy, A. C.

- (2022). Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nature methods*, 19, 429–440.
- Mi, F., Hu, C., Wang, Y., et al. (2022). Recent advancements in microfluidic chip biosensor detection of foodborne pathogenic bacteria: a review. *Anal Bioanal Chem* 414, 2883–2902.
- Microbiome. Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/microbiome>. Accessed 25 Jan. 2023.
- Milani, C., Duranti, S., Bottacini, F., Casey, E., Turrone, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arbolea Montes, S., Mancabelli, L., Lugli, G. A., Rodriguez, J. M., Bode, L., de Vos, W., Gueimonde, M., Margolles, A., van Sinderen, D., & Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and molecular biology reviews: MMBR*, 81(4), e00036-17.
- Miller, T. L., & Wolin, M. J. (1981). Fermentation by the human large intestine microbial community in an in vitro semicontinuous culture system. *Applied and Environmental Microbiology*, 42(3), 400–407.
- Mirzaee, M., Song, M., & Charmchi, H. S. (2016). Microfluidics based on-chip impinger for airborne particle collection. *Lab on a chip*, 16.
- Miyamoto, M., Motooka, D., & Gotoh, K., et al. (2014) Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics* 15, 699.
- Moelling, K., & Broecker, F. (2020). Air Microbiome and Pollution: Composition and Potential Effects on Human Health, Including SARS Coronavirus Infection. *Journal of environmental and public health*, 1646943.
- Moffatt, M. F., & Cookson, W. O. (2017). The lung microbiome in health and disease. *Clinical medicine (London, England)*, 17(6), 525–529.
- Moffitt, J. R., Lee, J. B., & Cluzel, P. (2012). The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab on a Chip*, 12(8), 1487–1494.
- Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., & Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J*, 4, 673–685.
- Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., & Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, 420, 806–810.

- Müller, T., & Winter, D. (2017). Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Molecular & cellular proteomics*, 16(7), 1173–1187.
- Nakagawa, T., Tanaka, T., Niwa, D., Osaka, T., Takeyama, H., & Matsunaga, T. (2005). Fabrication of amino silane-coated microchip for DNA extraction from whole blood. *J. Biotechnol*, 116, 105 – 111.
- Nan, L., Jiang, Z., & Wei, X. (2014). Emerging microfluidic devices for cell lysis: A review. *Lab Chip*, 14, 83.
- Ndiaye, M. M., Ta, H. P., Chiappetta, G., & Vinh, J. (2020). On-Chip Sample Preparation Using a ChipFilter Coupled to NanoLC-MS/MS for Bottom-Up Proteomics. *Journal of Proteome Research*, 19(7), 2654–2663.
- Ng, C., DeMaere, M.Z., Williams, T.J., Lauro, F.M., Raftery, M., Gibson, J.A., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J.M., Thomas, T., et al. (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *ISME J*, 4, 1002–1019.
- Nguyen, T. L. A., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease Models & Mechanisms*, 8(1), 1–16.
- Nie, K., Ma, K., Luo, W., Shen, Z., Yang, Z., Xiao, M., Tong, T., Yang, Y., & Wang, X. (2021). *Roseburia intestinalis*: A Beneficial Gut Organism From the Discoveries in Genus and Species. *Frontiers in Cellular and Infection Microbiology*. 11.
- Nilsson, J., Evander, M., Hammarström, B., & Laurell, T. (2009). Review of cell and particle trapping in microfluidic systems. *Analytica Chimica Acta*, 649(2), 141–157.
- Oh, H. J., Park, J. Y., Park, S. E., Lee, B. Y., Park, J. S., Kim, S.-K., Yoon, T. J., & Lee, S.-H. (2009). DNA-Enrichment Microfluidic Chip for Chromatin Immunoprecipitation. *Analytical Chemistry*, 81(8), 2832–2839.
- Orozco-Mosqueda, M., Rocha-Granados, M., Glick, B. R., & Santoyo, G. (2018). Microbiome engineering to improve biocontrol and plant growth-promoting mechanisms. *Microbiol Res*, 208, 5–31.
- Osorio, D., Rondón-Villarreal, P., & Torres, R. (2015). Peptides: A Package for Data Mining of Antimicrobial Peptides. *R Journal*, 7 (1), 4.
- Ota, M., & Fujio, K. (2021). Multi-omics approach to precision medicine for immune-mediated diseases. *Inflammation and Regeneration*, 41(1), 23.
- Ottesen, E. A., Hong, J. W., Quake, S. R., & Leadbetter, J. R. (2006). Microfluidic Digital PCR Enables Multigene Analysis of Individual Environmental Bacteria. *Science*, 314(5804), 1464–1467.

- Pammi, M., Aghaeepour, N., & Neu, J. (2022). Multiomics, artificial intelligence, and precision medicine in perinatology. *Pediatric Research*.
- Pan, S., & Chen, R. (2020). Metaproteomic analysis of human gut microbiome in digestive and metabolic diseases. *Advances in clinical chemistry*, 97, 1–12.
- Parapouli, M., Vasileiadis, A., Afendra, A. S., & Hatziloukas, E. (2020). *Saccharomyces cerevisiae* and its industrial applications. *AIMS microbiology*, 6(1), 1–31.
- Park, D. S.-W., Hupert, M.L., Witek, M.A., You, B.H., Datta, P., Guy, J., Lee, J.-B., Soper, S.A., Nikitopoulos, D.E., & Murphy, M.C. (2008). A titer plate-based polymer microfluidic platform for high throughput nucleic acid purification. *Biomed. Microdevices*, 10, 21 – 33.
- Park, J., Kerner, A., Burns, M. A., & Lin, X. N. (2011). Microdroplet-Enabled Highly Parallel Co-Cultivation of Microbial Communities. *PLoS ONE*, 6(2), e17019.
- Perez-Carrasco, V., Soriano-Lerma, A., Soriano, M., Gutiérrez-Fernández, J., & Garcia-Salcedo, J. A. (2021). Urinary Microbiome: Yin and Yang of the Urinary Tract. *Frontiers in Cellular and Infection Microbiology*, 11.
- Pérez-Cobas, A. E., Gomez-Valero, L., & Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial genomics*, 6, mgen000409.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., ... Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, 47(D1), D442–D450.
- Pérez-Rodríguez, S., Ramírez, O. T., Trujillo-Roldán, M. A., & Valdez-Cruz, N. A. (2020). Comparison of protein precipitation methods for sample preparation prior to proteomic analysis of Chinese hamster ovary cell homogenates. *Electronic Journal of Biotechnology*, 48, 86–94.
- Pfohl, T., Mugele, F., Seemann, R., & Herminghaus, S. (2003). Trends in microfluidics with complex fluids. *ChemPhysChem*, 4 (12), 1291–1298.
- Piendl, S. K., Geissler, D., Weigelt, L., & Belder, D. (2020). Multiple Heart-Cutting Two-Dimensional Chip-HPLC Combined with Deep-UV Fluorescence and Mass Spectrometric Detection. *Analytical Chemistry*, 92(5), 3795–3803.
- Ping, J., Kui, W., Canhua, H. & Edouard, C. N. (2017). Mining the fecal proteome: from biomarkers to personalised medicine, *Expert Review of Proteomics*.
- Piovesana, S., Capriotti, A. L., Foglia, P., Montone, C. M., La Barbera, G., Zenezini Chiozzi, R., Laganà, A., & Cavaliere, C. (2019). Development of an Analytical Method for the

Metaproteomic Investigation of Bioaerosol from Work Environments. *Proteomics*, 19(23), e1900152.

Plavchak, C. L., Smith, W. C., Bria, C. R. M., & Williams, S. K. R. (2021). New Advances and Applications in Field-Flow Fractionation. *Annual Review of Analytical Chemistry*, 14(1), 257–279.

Price, N. D., Magis, A. T., Earls, J. C., Glusman, G., Levy, R., Lausted, C., McDonald, D. T., Kusebauch, U., Moss, C. L., Zhou, Y., Qin, S., Moritz, R. L., Brogaard, K., Omenn, G. S., Lovejoy, J. C., & Hood, L. (2017). A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nature biotechnology*, 35(8), 747–756.

Psifidi, A., Dovas, C.I., Bramis, G., et al. (2015). Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. *PLoS One*, 10, 0115960.

Qin, N., Zhao, P., Ho, E. A., Xin, G., & Ren, C. L. (2021). Microfluidic Technology for Antibacterial Resistance Study and Antibiotic Susceptibility Testing: Review and Perspective. *ACS Sensors*, 6(1), 3–21.

Rajasekaran, S., Siddiqui, J., Rakijas, J., Nicolay, B., Lin, C., Khan, E., Patel, R., Morris, R., Wyler, E., Boukhali, M., Balasubramanyam, J., Ranjith Kumar, R., Van Rechem, C., Vogel, C., Elchuri, S. V, Landthaler, M., Obermayer, B., Haas, W., Dyson, N., & Miles, W. (2021). Integrated multi-omics analysis of RB-loss identifies widespread cellular programming and synthetic weaknesses. *Communications Biology*, 4(1), 977.

Rakszewska, A., Tel, J., Chokkalingam, V., et al. (2014). One drop at a time: toward droplet microfluidics as a versatile tool for single-cell analysis. *NPG Asia Mater*, 6, e133.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah, M., Hettich, R. L., & Banfield, J. F. (2005). Community proteomics of a natural microbial biofilm. *Science*, 308, 1915–1920.

Rauniyar, N. (2015). Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *International journal of molecular sciences*, 16(12), 28566–28581.

Raza, S., Jo, H., Kim, J., Shin, H., Hur, H.-G., & Unno, T. (2021). Metagenomic exploration of antibiotic resistome in treated wastewater effluents and their receiving water. *Science of The Total Environment*, 765, 142755.

Reinholt, S. J., & Baeumner, A. J. (2014). Microfluidic Isolation of Nucleic Acids. *Angewandte Chemie International Edition*, 53(51), 13988–14001.

Reinholt, S. J., Behrent, A., Greene, C., Kalfe, A., & Baeumner, A. J. (2014). A magnetic bead-based assay for the rapid detection of methicillin-resistant *Staphylococcus aureus* by

- using a microfluidic system with integrated loop-mediated isothermal amplification. *Anal. Chem.*, 86, 849 – 856.
- Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A., 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).
- Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C., & Gasbarrini, A. (2017). Proteobacteria: A Common Factor in Human Diseases. *BioMed research international*, 2017, 9351507.
- Robertson, S. J., Lemire, P., Maughan, H., et al. (2019). Comparison of cohousing and littermate methods for microbiota standardization in mouse models. *Cell Rep*, 27, 1910–1919.
- Rogers, Y.-H., & Zhang, C. (2016). Genomic Technologies in Medicine and Health: Past, Present, and Future. In: Kumar D, Antonarakis S, editors. *Med Health Genomics*. Oxford: Academic Press; 15–28.
- Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., et al. (2015). Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *npj Biofilms Microbiomes*, 1, 15007.
- Rusconi, R., Garren, M., & Stocker, R. (2014). Microfluidics Expanding the Frontiers of Microbial Ecology. *Annual Review of Biophysics*, 43(1), 65–91.
- Russo, D. A., Couto, N., Beckerman, A. P., & Pandhal, J. (2016). A Metaproteomic Analysis of the Response of a Freshwater Microbial Community under Nutrient Enrichment. *Frontiers in Microbiology*, 7.
- Ryan, K.J. & Ray, C.G., Eds. (2004). *Sherris Medical Microbiology: an introduction to Infectious Diseases*. 4th Edition, McGraw-Hill, New York, page 310. ISBN 978-0-8385-8529-0.
- Sahoo, S., Ravi Kumar, R.K., Nicolay, B., Mohite, O., Sivaraman, K., Khetan, V., Rishi, P., Ganesan, S., Subramanyan, K., Raman, K., Miles, W. & Elchuri, S.V. (2019). Metabolite systems profiling identifies exploitable weaknesses in retinoblastoma. *FEBS Lett*, 593: 23-41.
- Salvato, F., Hettich, R. L., & Kleiner, M. (2021). Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLoS Pathog*, 17(2), e1009245.
- Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K., & Gordon, J. I. (2007). Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proceedings of the National Academy of Sciences*, 104(25), 10643–10648.

- Samuel, B., Zhou, L., Jennifer, P.-R., L., H. R., Xavier, M., Chongle, P., & Ryan, S. M. (2016). Proteomic Stable Isotope Probing Reveals Taxonomically Distinct Patterns in Amino Acid Assimilation by Coastal Marine Bacterioplankton. *MSystems*, 1(2), e00027-15.
- Šantl-Temkiv, T., Amato, P., Casamayor, E. O., Lee, P. K. H., & Pointing, S. B. (2022). Microbial ecology of the atmosphere. *FEMS Microbiology Reviews*, 46, fuac009.
- Sartaj, K., Patel, A., Matsakas, L., & Prasad, R. (2022). Unravelling Metagenomics Approach for Microbial Biofuel Production. *Genes*, 13, 1942.
- Schlaberg, R. (2020). Microbiome Diagnostics. *Clinical chemistry*, 66(1), 68–76.
- Schlaeppli, K., & Bulgarelli, D. (2015). The plant microbiome at work. *Mol Plant-Microbe Interact*, 28, 212–7.
- Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., Richter, A., Eberl, L., Zechmeister-Boltenstern, S., & Riedel, K. (2012). Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J*, 6, 1749–1762.
- Scott, S. M., & Ali, Z. (2021). Fabrication Methods for Microfluidic Devices: An Overview. *Micromachines*, 12(3), 319.
- Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., DeMaere, M. Z., ... McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation-a benchmark of metagenomics software. *Nature methods*, 14, 1063–1071.
- Sender, R., Fuchs, S., & Milo, R. (2016.A). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*, 14(8): e1002533.
- Sender, R., Fuchs, S., & Milo, R. (2016.B). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, 164, 337–340.
- Shaffer, J. P., Nothias, L.-F., Thompson, L. R., Sanders, J. G., Salido, R. A., Couvillion, S. P., Brejnrod, A. D., Lejzerowicz, F., Haiminen, N., Huang, S., Lutz, H. L., Zhu, Q., Martino, C., Morton, J. T., Karthikeyan, S., Nothias-Esposito, M., Dührkop, K., Böcker, S., Kim, H. W., ... Consortium, the E. M. P. 500 (EMP500). (2022). Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. *Nature Microbiology*, 7(12), 2128–2150.
- Sharma, P., Kumar, S., & Pandey, A. (2021). Bioremediated techniques for remediation of metal pollutants using metagenomics approaches: A review. *Journal of Environmental Chemical Engineering*, 9, 105684.

- Shemesh, J., Bransky, A., Khoury, M., & Levenberg, S. (2010). Advanced microfluidic droplet manipulation based on piezoelectric actuation. *Biomedical microdevices*, 12(5), 907–914.
- Shen, F., Tan, M., Wang, Z., Yao, M., Xu, Z., Wu, Y., Wang, J., Guo, X., & Zhu, T. (2011). Integrating silicon nanowire field effect transistor, microfluidics and air sampling techniques for real-time monitoring biological aerosols. *Environmental science & technology*, 45 (17), 7473-80.
- Shilton, R. J., Travagliati, M., Beltram, F., & Cecchini, M. (2014). Nanoliter-droplet acoustic streaming via ultra high frequency surface acoustic waves. *Advanced materials (Deerfield Beach, Fla.)*, 26(29), 4941–4946.
- Signat, B., Roques, C., Poulet, P., & Duffaut, D. (2011). *Fusobacterium nucleatum* in periodontal health and disease. *Current issues in molecular biology*, 13(2), 25–36.
- Singleton, I., Merrington, G., Colvan, S., & Delahunty, J. S. (2003). The potential of soil protein-based methods to indicate metal contamination. *Appl. Soil Ecol*, 23, 25–32.
- Skenneron, C. T., Barr, J. J., Slater, F. R., Bond, P. L., & Tyson, G. W. (2015). Metabolic plasticity in *Accumulibacter* clades. *Environ Microbiol*, 17, 1574-1585.
- Smets, W., Moretti, S., Denys, S., & Lebeer, S. (2016). Airborne bacteria in the atmosphere: Presence, purpose, and potential. *Atmospheric Environment*, 139, 214–221.
- Sowell, S. M., Abraham, P. E., Shah, M., Verberkmoes, N. C., Smith, D. P., Barofsky, D. F., & Giovannoni, S. J. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J*, 5, 856–865.
- Speth, D., in 't Zandt, M., Guerrero-Cruz, S., et al. (2016). Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat Commun*, 7, 11172.
- Stewart, J. D., Kremer, P., Shakya, K. M., Conway, M., & Saad, A. (2021). Outdoor Atmospheric Microbial Diversity Is Associated With Urban Landscape Structure and Differs From Indoor-Transit Systems as Revealed by Mobile Monitoring and Three-Dimensional Spatial Analysis. *Frontiers in Ecology and Evolution*, 9.
- Streets, A. M., Zhang, X., Cao, C., Pang, Y., Wu, X., Xiong, L., Yang, L., Fu, Y., Zhao, L., Tang, F., & Huang, Y. (2014). Microfluidic single-cell whole-transcriptome sequencing. *Proceedings of the National Academy of Sciences*, 111(19), 7048–7053.
- Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*, 14, 1177932219899051.

- Swarge, B. N., Roseboom, W., Zheng, L., Abhyankar, W. R., Brul, S., de Koster, C. G., & de Koning, L. J. (2018). "One-Pot" Sample Processing Method for Proteome-Wide Analysis of Microbial Cells and Spores. *PROTEOMICS – Clinical Applications*, 12(5), 1700169.
- Tanca, A., Biossa, G., Pagnozzi, D., Addis, M. F., & Uzzau, S. (2013). Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the *Escherichia coli* proteome. *Proteomics*, 13 (17), 2597–2607.
- Tang, Q., Jin, G., Wang, G., Liu, T., Liu, X., Wang, B., & Cao, H. (2020). Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices, *Frontiers in Cellular and Infection Microbiology*, 10.
- Tewhey, R., Warner, J., Nakano, M., et al. (2009). Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol*, 27, 1025–1031.
- Thaitrong, N., Kim, H., Renzi, R. F., Bartsch, M. S., Meagher, R. J. & Patel, K. D. (2012). Quality control of next-generation sequencing library through an integrative digital microfluidic platform. *Electrophoresis*, 33, 3506-3513.
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp*, 2, 3.
- Todar, K. (2006). The Normal Bacterial Flora of Humans. *Todar's Online Textbook of Bacteriology*. University of Wisconsin-Madison Department of Bacteriology.
- Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and nextgeneration sequencing: computational challenges and solutions. *Nat Rev Genet*, 13, 36–46.
- Truckenmüller, R., Rummler, Z., Schaller, T., & Schomburg, W. K. (2002). Low-cost thermoforming of micro fluidic analysis chips. *Journal of Micromechanics and Microengineering*, 12 (4), 375–379.
- Tseng, Q., Lomonosov, A. M., Furlong, E. E. M., & Merten, C. A. (2012). Fragmentation of DNA in a sub-microliter microfluidic sonication device. *Lab on a Chip*, 12(22), 4677–4682.
- Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*, 11, 2301–2319.
- Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ*, 361, k2179.
- Van Den Bossche, T., Kunath, B.J., Schallert, K., et al. (2021). Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. *Nat Commun*. 12, 7305.
- Vandeputte, D., Kathagen, G., D’Hoe, K., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, 551, 507–511.

- Varnavides, G., Madern, M., Anrather, D., Hartl, N., Reiter, W., & Hartl, M. (2022). In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *Journal of Proteome Research*, 21(10), 2397–2411.
- Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., Carlsson, C. M., Asthana, S., Zetterberg, H., Blennow, K., et al. (2017). Gut microbiome alterations in Alzheimer's disease. *Sci. Rep.*, 7, 13537.
- Vulto, P., Dame, G., Maier, U., Makohliso, S., Podszun, S., Zahn, P., & Urban, G. A. (2010). A microfluidic approach for high efficiency extraction of low molecular weight RNA. *Lab Chip*, 10, 610 – 616.
- Waldner, J. B. (2010). *Nanocomputers and Swarm Intelligence*. London: ISTE John Wiley & Sons. p. 93. ISBN 978-1-84704-002-2.
- Wang, C.-H., Lien, K.-Y., Wu, J.-J., & Lee, G.-B. (2011.A). A magnetic bead-based assay for the rapid detection of methicillin-resistant *Staphylococcus aureus* by using a microfluidic system with integrated loop-mediated isothermal amplification *Lab Chip*, 11, 1521 – 1531.
- Wang, D.-Z., Dong, H.-P., Xie, Z.-X., Dai, M.-H., & Hong, H.-S. (2011.B). Metaproteomic characterization of dissolved organic matter in the water column of the South China Sea. *Limnol. Oceanogr*, 56, 1641–1652.
- Wang, H., Chen, L. & Sun, L. (2017). Digital microfluidics: A promising technique for biochemical applications. *Front. Mech. Eng.* 12, 510–525.
- Wang, J., Lang, T., Shen, J., Dai, J., Tian, L., & Wang, X. (2019). Core Gut Bacteria Analysis of Healthy Mice. *Frontiers in Microbiology*, 10.
- Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., Jia, W., Cai, S., & Zhao, L. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J*, 6, 320–329.
- Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical microbiology reviews*, 20(4), 593–621.
- Whipps, J., Lewis, K., & Cooke, R. (1988). Mycoparasitism and plant disease control. In: Burge M, editor. *Fungi Biol Control Syst*. Manchester University Press, 161-187.
- White side, S. A., Razvi, H., Dave, S., Reid, G., & Burton, J. P. (2015). The microbiome of the urinary tract—a role beyond infection. *Nat Rev Urol*, 12(2), 81.
- White, F. M. (1991). *Viscous Fluid Flow*. McGraw-Hill.
- Wiebelhaus, N., Zaengle-Barone, J. M., Hwang, K. K., Franz, K. J., & Fitzgerald, M. C. (2021). Protein Folding Stability Changes Across the Proteome Reveal Targets of Cu Toxicity in *E. coli*. *ACS chemical biology*, 16(1), 214–224.

- Wilkins, M. J., Verberkmoes, N. C., Williams, K. H., Callister, S. J., Mouser, P. J., Elifantz, H., N'Guessan A, L., Thomas, B. C., Nicora, C. D., Shah, M. B., et al. (2009). Proteogenomic monitoring of geobacter physiology during stimulated uranium bioremediation. *Appl. Environ. Microbiol*, 75, 6591–6599.
- Williams, T. J., Long, E., Evans, F., Demaere, M. Z., Lauro, F. M., Raftery, M. J., Ducklow, H., Grzymiski, J. J., Murray, A. E., & Cavicchioli, R. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J*, 6, 1883–1900.
- Wilmes, P., & Bond, P. L. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol*, 14, 92-97.
- Wilmes, P., Andersson, A. F. Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L., Bond, P. L., VerBerkmoes, N. C., & Banfield, J. F. (2008.A). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J*, 2, 853–864.
- Wilmes, P., Heintz-Buschart, A. & Bond, P.L. (2015). A decade of metaproteomics: Where we stand and what the future holds. *Proteomics*, 15, 3409-3417.
- Wilmes, P., Wexler, M., & Bond, P.L. (2008.B). Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS ONE*, 3, 1778.
- Wiśniewski, J. R., & Rakus, D. (2014). Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. *Journal of Proteomics*, 109, 322–331.
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, 6(5), 359–362.
- Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), 5088–5090.
- Wook Hong, J., Chen, Y., French Anderson, W., & Quake, S. R. (2006). Molecular biology on a microfluidic chip. *Journal of Physics: Condensed Matter*, 18(18), S691.
- Wu, A. R., Hiatt, J. B., Lu, R., Attema, J. L., Lobo, N. A., Weissman, I. L., Clarke, M. F., & Quake, S. R. (2009). Automated microfluidic chromatin immunoprecipitation from 2,000 cells. *Lab on a Chip*, 9(10), 1365–1370.
- Wu, J., Zhu, J., Yin, H., Liu, X., An, M., Pudlo, N. A., Martens, E. C., Chen, G. Y., & Lubman, D. M. (2016). Development of an Integrated Pipeline for Profiling Microbial Proteins from Mouse Fecal Samples by LC–MS/MS. *Journal of Proteome Research*, 15(10), 3635–3642.

- Xiao, L., Feng, Q., Liang, S., Sonne, S. B., Xia, Z., Qiu, X., Li, X., Long, H., Zhang, J., Zhang, D., Liu, C., Fang, Z., Chou, J., Glanville, J., Hao, Q., Kotowska, D., Colding, C., Licht, T. R., Wu, D., ... Kristiansen, K. (2015). A catalog of the mouse gut metagenome. *Nature Biotechnology*, 33(10), 1103–1108.
- Xie, J., He, Z., Liu, X., Liu, X., van Nostrand, J.D., Deng, Y., Wu, L., Zhou, J., & Qiu, G. (2011). Geochip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. *Appl. Environ. Microbiol*, 77, 991–999.
- Xu, L., Brito, I., Alm, E., et al. (2016). Virtual microfluidics for digital quantification and single-cell sequencing. *Nat Methods*, 13, 759–762.
- Xue, W., Li, J. J., Zou, Y., Zou, B., & Wei, L. (2021). Microbiota and Ocular Diseases. *Frontiers in cellular and infection microbiology*, 11, 759333.
- Yadav, M., & Chauhan, N. S. (2021). Microbiome therapeutics: exploring the present scenario and challenges. *Gastroenterology report*, 10, goab046.24
- Yawata, Y., Nguyen, J., Stocker, R., & Rusconi, R. (2016). Microfluidic studies of biofilm formation in dynamic environments. *Journal of Bacteriology*, 198(19), 2589–2595.
- Yong, S. C., Roversi, P., Lillington, J., Rodriguez, F., Krehenbrink, M., Zeldin, O. B., Garman, E. F., Lea, S. M., & Berks, B. C. (2014). A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. *Science*, 345, 1170–1173.
- Young, E. W., Berthier, E., Guckenberger, D. J., Sackmann, E., Lamers, C., Meyvantsson, I., et al. (2011). Rapid prototyping of arrayed microfluidic systems in polystyrene for cell-based assays. *Analytical Chemistry*, 83 (4), 1408–1417.
- Yu, Z., Lu, S., & Huang, Y. (2014). Microfluidic Whole Genome Amplification Device for Single Cell Sequencing. *Analytical Chemistry*, 86(19), 9386–9390.
- Yue, G. E., Roper, M. G., Balchunas, C., Pulsipher, A., Coon, J. J., Shabanowitz, J., Hunt, D. F., Landers, J. P., & Ferrance, J. P. (2006). Protein digestion and phosphopeptide enrichment on a glass microchip. *Analytica Chimica Acta*, 564(1), 116–122.
- Zhang, R., Gong, H.-Q., Zeng, X., Lou, C., & Sze, C. (2013). A Microfluidic Liquid Phase Nucleic Acid Purification Chip to Selectively Isolate DNA or RNA from Low Copy/Single Bacterial Cells in Minute Sample Volume Followed by Direct On-Chip Quantitative PCR Assay. *Anal. Chem.*, 85, 1484 – 1491.
- Zhang, X., Deeke, S.A., Ning, Z., et al. (2018). Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun*, 9, 2873.
- Zhang, Y., & Nguyen, N. T. (2017). Magnetic digital microfluidics - a review. *Lab on a chip*, 17(6), 994–1008.

Zhou, P., Manoil, D., Belibasakis, G. N., & Kotsakis, G. A. (2021). Veillonellae: Beyond Bridging Species in Oral Biofilm Ecology. *Frontiers in Oral Health*. 2.

RÉSUMÉ

L'étude des microbiomes répond à trois grandes questions : (i) qui sont-ils ? (ii) que peuvent-ils faire ? (iii) que font-ils ? Deux technologies analytiques - la spectrométrie de masse et le séquençage de nouvelle génération - ont joué un rôle essentiel dans l'étude du microbiome et sont couramment utilisées pour le séquençage des protéines et des acides nucléiques respectivement. Les flux de travail existants pour le traitement pré-analytique des échantillons microbiens sont limités car ils ne sont pas intégrés pour effectuer la préparation des échantillons pour les protéines et les acides nucléiques.

L'étude actuelle démontre de nouveaux flux de travail utilisant un dispositif microfluidique appelé ChipFilter pour effectuer une préparation d'échantillon automatisée et intégrée pour la métaprotéomique avec l'isolement séquentiel des acides nucléiques. Tous les flux de travail nécessaires aux processus mentionnés ont été démontrés avec des mélanges microbiens standard de trois espèces à nombre de cellules égal et de dix-sept espèces à nombre de cellules disproportionné. En outre, la validation du flux de travail métaprotéomique a été effectuée en utilisant le microbiome intestinal de la souris isolé à partir de ses fèces. Les résultats montrent des identifications élevées de protéines et une isolation stable des acides nucléiques confirmée par amplification PCR. En conclusion, ChipFilter a été créé avec le potentiel d'être facilement intégré pour la collecte d'échantillons et la préparation d'échantillons multi omiques pour les études de microbiome.

MOTS CLÉS

Spectrometrie de Masse ; Microfluidique ; Metaproteomique

ABSTRACT

Studying microbiomes answers three broad question (i) who is there? (ii) what can they do? (iii) what are they doing? Two analytical technologies – mass spectrometry and next generation sequencing, has been vital for the microbiome studies and are routinely used for sequencing proteins and nucleic acids respectively. Existing workflows for pre-analytical sample processing of microbial samples have been limited by not being integrated to perform sample preparation for both protein and nucleic acids.

The current study will demonstrate novel workflows utilizing a microfluidic device named ChipFilter to perform automated and integrated sample preparation for metaproteomics with sequential nucleic acid isolation. All the required workflows to the mentioned processes have been demonstrated with standard microbial mixtures of three equal cell number and seventeen disproportionate cell number species. Additionally, metaproteomics workflow validation was done using the mouse gut microbiome isolated from mice faeces. Results show high protein identifications and stable nucleic acid isolation confirmed by PCR amplication. In conclusion, ChipFilter with potential to be readily integrated for sample collection and performing multi omics sample preparation for microbiome studies was established.

KEYWORDS

Mass Spectrometry ; Microfluidics ; Metaproteomics

