COMPARAISON DE DEUX OUTILS BIOINFORMATIQUES POUR L'ETUDE DE LA BIODIVERSITE DES BIOAEROSOLS DANS LES LIEUX DE TRAVAIL

L. Alonso*[1], J. Degois[1] and P. Duquenne[2]

[1] Laboratoire de Métrologie des Aérosols
[2] Laboratoire d'Analyses Spatiales et Temporelles des Expositions Chimiques
INRS, 54519 Vandœuvre-lès-Nancy Cedex, France
*Courriel de l'orateur : lise.alonso@inrs.fr

TITLE

Comparison of two bioinformatics tools to study biodiversity of bioaerosols in the workplace

RESUME

Connaître la biodiversité des bactéries et des champignons dans les atmosphères de travail est devenu essentiel pour mieux appréhender l'exposition aux bioaérosols et réduire le risque d'exposition. La mise en culture est la méthode la plus utilisée pour identifier et quantifier les bioaérosols, mais n'identifie que 1 à 10% des micro-organismes. D'autres méthodes basées sur la biologie moléculaire, comme le séquençage à haut débit (HTS), permettent d'étudier la quasi-totalité des microorganismes d'un échantillon d'aérosol. Le HTS nécessite une analyse bioinformatique pour identifier et quantifier la communauté microbienne. Plusieurs outils bioinformatiques existent, mais aucune information n'est disponible concernant leurs similitudes ou leurs différences. Dans ce travail, nous avons comparé deux outils bioinformatiques largement utilisés, Mothur et FROGS, sur des échantillons d'aérosols d'une usine de tri de déchets. Les deux outils apportent des résultats similaires pour l'identification des phylums les plus abondants. Ce travail mérite d'être approfondi au niveau du genre et pourquoi pas en intégrant d'autres outils de bioinformatique.

ABSTRACT

Knowing about biodiversity of bacteria and fungi in workplace atmospheres has become essential towards being able to amin the risk of exposure. Culture is the most method used to identify and quantify bioaerosols but only identifies 1 to 10% of microorganisms. Others method based on molecular biology like high-throughput sequencing (HTS) make it possible to study almost all microorganisms in an aerosol sample. The HTS required bioinformatics analysis to identify and quantify microbial community. Several bioinformatics tools existed, but no information is available concerning their similarities or their differences. In this work, we compared two widely used bioinformatics tools, Mothur and FROGS, on aerosols samples of waste sorting plant. Both tools presented a high degree of agreement at identifying the most abundant phylums. This work deserves to be deepened at the genus level and why not by integrating other bioinformatics tools.

MOTS-CLÉS : bioaérosols, séquençage haut débit, bioinformatique, identification / **KEYWORDS:** bioaerosols, high-throughput sequencing, bioinformatics tools, identification

1. CONTEXTE ET OBJECTIF

Une préoccupation croissante concerne les bioaérosols et leurs effets nocifs sur la santé. Connaître la biodiversité des bactéries et des champignons dans les atmosphères de travail est devenu une étape importante pour mieux appréhender l'exposition aux bioaérosols et réduire le risque d'exposition. La culture est la méthode la plus utilisée pour identifier et quantifier les bioaérosols dans les atmosphères de travail. La culture est simple d'utilisation et peu coûteuse, mais n'identifie que 1 à 10% des microorganismes (Duquenne, 2018), donc seulement une petite partie de la communauté microbienne. D'autres méthodes basées sur la biologie moléculaire comme le séquençage haut débit permettent d'étudier la quasi-totalité des microorganismes d'un échantillon d'aérosol. Le séquençage haut débit nécessite une analyse bioinformatique pour identifier et quantifier la communauté microbienne. A ce jour, plusieurs outils de traitement de données bioinformatiques sont disponibles mais n'ont pas été comparés pour l'analyse des bioaérosols en milieu de travail. Ainsi, aucune information n'est disponible concernant les similitudes ou les différences qui pourraient exister entre ces outils.

Dans ce travail, nous avons utilisé deux outils bioinformatiques largement répandu, Mothur (Schloss et al.2009) et FROGS (Escudié et al.2017). Mothur et FROGS sont des logiciels en libre accès, Mothur est un programme en ligne de commande qui nécessite une certaine connaissance du code alors que FROGS est convivial et conçu pour les non-spécialistes. Le but étant de comparer les résultats issus du traitement de données des séquences et d'évaluer les similitudes ou différences obtenues.

DOI : 10.25576/ASFERA-CFA2021-24835 Cet article est publié sous la seule responsabilité de ses auteurs

2. MATERIELS ET METHODES

Cette comparaison a été réalisée sur des séquences issues d'échantillons de bioaérosols prélevés pendant six mois (janvier à juin) dans une usine de tri de déchets en France (Degois, 2018). Ces échantillons ont été prélevés par filtration à différents endroits en utilisant des cassettes fermées en polypropylène et constituées de 3 pièces (Millipore®, France). Pour le prélèvement, les cassettes fermées étaient chacune connectées à une pompe individuelle (SG 10-2, GSA) assurant un débit nominal moyen de 10 L/min. Pour étudier la diversité de la communauté microbienne de façon globale, la méthode de séquençage haut débit Illumina a été choisie. Deux marqueurs génétiques de diversité microbienne ont été étudiés : les gènes ARNr 16S spécifiques aux bactéries et *l'Internal Transcribed Spacer* 1 (ITS1) spécifiques aux champignons. Pour les amplifications du gène de l'ARNr 16S, les amorces ciblant les régions V4-V5 ont été utilisées. Le séquençage haut débit a été réalisé par INRAE transfert (Narbonne), en utilisant la chimie appariée Illumina MiSeq 2 x 250 pb pour obtenir entre 30 000 et 100 000 séquences par échantillon.

Le traitement de données des séquences obtenues par le séquençage haut débit a été effectué à l'aide des outils FROGS et Mothur. Le processus comprend plusieurs étapes qui peuvent varier en fonction de l'outil utilisé.
Avec FROGS, les séquences sont assemblées avec un maximum de 10% de discordance dans la région superposée, puis les séquences n'ayant pas la longueur attendue (c'est-à-dire 200 à 500 pb) ou contenant des bases ambiguës (N) ont été supprimées. Ensuite, les séquences sont regroupées grâce à l'outil SWARM, en fonction de leur degré de similarité afin de former un groupe ou cluster associé à un OTU (Unité opérationnelle taxonomique). Les séquences chimériques ont ensuite été éliminées en utilisant VSEARCH et les séquences de faible abondance ont été filtrées à ≥ 0,005%, c'est-à-dire en conservant les OTU représentant au moins 0,005% de toutes les séquences. Pour finir, l'affiliation taxonomique a été réalisée avec la base de données SILVA pour les bactéries et la base de données ITS UNITE pour les champignons.
Avec Mothur, les séquences sont appariées et dérépliquées, c'est-à-dire qu'un seul exemplaire de toutes les séquences identiques est conservé. Les séquences contenant des bases ambiguës (N) sont supprimées. Puis, les séquences sont regroupées avec un degré de similarité de 97%. Ensuite le traitement de données est similaire à celui effectué avec FROGS, c'est-à-dire que les séquences chimériques sont supprimées et l'affiliation taxonomique est réalisé à l'aide des mêmes bases de données.
Ce processus abouti pour les deux outils à une quantification relative et à l'identification de la communauté bactérienne et fongique au niveau du phylum.

3. RESULTATS ET DISCUSSION

La composition de la communauté bactérienne au niveau du phylum obtenue par les deux outils est proche et permet d'identifier les phylums les plus abondants (Figure 1) : Actinobacteria, Firmicutes, Proteobacteria, Bacteroidetes et Cyanobacteria. Cependant, des différences sont constatées en avril et mai avec une proportion plus élevée de Protéobactéries pour FROGS (66% en avril et 68% en mai) que Mothur (24% en avril et 29% en mai). Cette différence pourrait s'expliquer par la méthode de regroupement utilisée, en effet un seuil local (distance d'agrégation de 3) est appliqué avec FROGS alors qu'un seuil global (3%) est appliqué avec Mothur. Des résultats similaires ont été obtenus avec la communauté fongique (données non présentées).

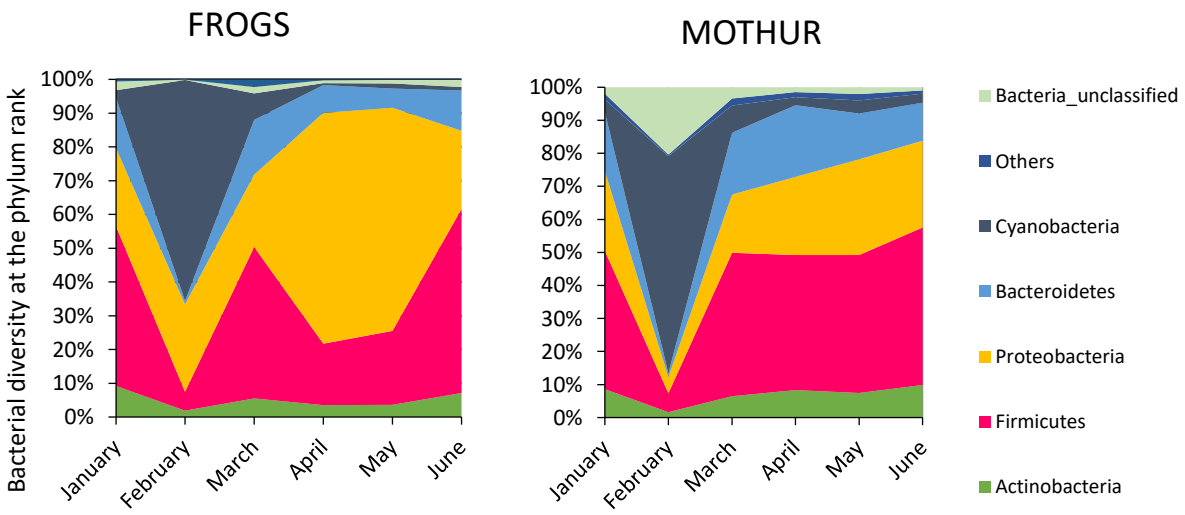

Figure 1. Composition en phylum bactériens des bioaérosols identifiés par FROGS et Mothur

DOI : 10.25576/ASFERA-CFA2021-24835
Cet article est publié sous la seule responsabilité de ses auteurs

4. CONCLUSION

Dans ce travail, les outils de traitement de données de la biodiversité des aérosols, FROGS et Mothur présentent des résultats similaires. Il faut souligner que l'objectif n'était pas de déterminer quel outil fournit une image plus précise de la véritable composition des bioaérosols, mais d'être conscient des différences et limites potentielles entre les différents outils disponibles. Cette étude comparative n'a été réalisé qu'au niveau du phylum, il serait intéressant d'étudier des rangs taxonomiques plus bas afin de comparer plus finement ces deux outils, mais aussi d'inclure un troisième outil bioinformatique très répandu Qiime.

Degois, J. (2018). *Diversité microbienne dans les bioaérosols émis dans les centres de tri des déchets* (thèse de doctorat). Université de Lorraine, France

Duquenne, P. (2018) *Ann. Work Expo. Health*, 62, 139-146.

Escudié, F. *et al.* (2017) *Bioinformatics*, 34, 1287-94.

Schloss, P.D. *et al*. (2009) *Appl. and Environ. Microbiol*, 75, 7537-41.

DOI : 10.25576/ASFERA-CFA2021-24835 Cet article est publié sous la seule responsabilité de ses auteurs