



Article

Classification of Environmental Strains from Order to Genus Levels Using Lipid and Protein MALDI-ToF Fingerprintings and Chemotaxonomic Network Analysis

Marceau Levasseur¹, Téo Hebra¹, Nicolas Elie¹ , Vincent Guérineau¹, David Touboul^{1,2,*} and Véronique Eparvier^{1,*}

- ¹ CNRS, Institut de Chimie des Substances Naturelles (ICSN), UPR 2301, Université Paris-Saclay, Avenue de la Terrasse, 91 198 Gif-sur-Yvette, France; marceau.levasseur@cnrs.fr (M.L.); teo.hebra@cnrs.fr (T.H.); nicolas.elie@cnrs.fr (N.E.); vincent.guerineau@cnrs.fr (V.G.)
- ² Laboratoire de Chimie Moléculaire (LCM), CNRS UMR 9168, École Polytechnique, Institut Polytechnique de Paris, Route de Saclay, CEDEX, 91 128 Palaiseau, France
- * Correspondence: david.touboul@cnrs.fr (D.T.); veronique.eparvier@cnrs.fr (V.E.); Tel.: +33-169-823-032 (D.T.); +33-169-823-679 (V.E.)

Abstract: During the last two decades, MALDI-ToF mass spectrometry has become an efficient and widely-used tool for identifying clinical isolates. However, its use for classification and identification of environmental microorganisms remains limited by the lack of reference spectra in current databases. In addition, the interpretation of the classical dendrogram-based data representation is more difficult when the quantity of taxa or chemotaxa is larger, which implies problems of reproducibility between users. Here, we propose a workflow including a concurrent standardized protein and lipid extraction protocol as well as an analysis methodology using the reliable spectra comparison algorithm available in MetGem software. We first validated our method by comparing protein fingerprints of highly pathogenic bacteria from the Robert Koch Institute (RKI) open database and then implemented protein fingerprints of environmental isolates from French Guiana. We then applied our workflow for the classification of a set of protein and lipid fingerprints from environmental microorganisms and compared our results to classical genetic identifications using 16S and ITS region sequencing for bacteria and fungi, respectively. We demonstrated that our protocol allowed general classification at the order and genus level for bacteria whereas only the Botryosphaerales order can be finely classified for fungi.

Keywords: environmental microorganisms; classification; mass spectrometry; MALDI; molecular network



Citation: Levasseur, M.; Hebra, T.; Elie, N.; Guérineau, V.; Touboul, D.; Eparvier, V. Classification of Environmental Strains from Order to Genus Levels Using Lipid and Protein MALDI-ToF Fingerprintings and Chemotaxonomic Network Analysis. *Microorganisms* **2022**, *10*, 831. <https://doi.org/10.3390/microorganisms10040831>

Academic Editor: Adriana Calderaro

Received: 28 January 2022

Accepted: 13 April 2022

Published: 17 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Microorganisms are one of the first sources of enzymes and metabolites with wide applications in the industrial or health fields [1,2]. Their identification is necessary to access all the knowledge that has been shared by the scientific community thanks to the literature and open-source databases [3–5]. Since Woese's work in 1977, the identification of microorganisms has been based on molecular biology methods, in particular the sequencing of the gene coding for 16S rRNA, in the case of bacteria, or 18S rRNA or internal transcribed spacers (ITS) for fungi, and more recently, whole genome sequencing [6–8]. However, these methods still require careful sample preparation and the latter one has a high cost that does not favor its systematic application when dealing with large collections [6]. Furthermore, the taxonomic resolution offered by 16S rRNA gene sequencing is not sufficient to identify certain bacterial strains, unless third generation sequencers and denoising algorithms are used in combination [9]. As for fungi, ITS is not a universal barcode because of sequence redundancy between different species-rich genera [10].

Thus, the identification of microorganisms by MALDI-ToF mass spectrometry has become widely democratized during the last two decades; in the field of clinical microbiological diagnosis and agri-food industry, and finally for environmental studies [11–14]. MALDI-ToF mass spectrometry has proven to be a robust, rapid, and low-cost method for isolate identification or dereplication [15–18]. This is why one of the current methodologies used for human pathogenic strains consists in implementing private databases with new reference spectra. This method is experimentally and analytically tedious, in particular because the identification of redundant profiles is based on global similarity analyses by hierarchical clustering. Moreover, visually analyzing these dendrograms involves tedious human intervention due to the difficulty of reading and interpreting them as the amount of data analyzed increases [19].

Furthermore, MALDI-ToF mass spectrometry identification of representatives of a cultivable environmental microbial community proves to be more difficult than identification of a pathogenic microorganism due to the high proportion of only pathogenic microorganisms in the available databases [12,13]. Thus, fingerprint discrimination by detecting specificities could allow a better dereplication [20]. Among the chemo-informatic pipelines developed to allow classification and/or dereplication of metabolomics data, molecular networking is becoming more and more popular [21–24]. Its principle is based on the postulate that structurally close molecules share the same fragmentation pattern in mass spectrometry (MS). Although the construction of molecular networks has been used only for MS² data, it is also possible at MS¹ level, according to a direct comparison of MS¹ spectra [21,25,26]. Recently, Dumolin et al. demonstrated the robustness of this methodology through the creation of SPeDe; an open-source software for high-throughput dereplication of isolate fingerprints from MALDI-ToF MS analyses [27,28].

In this study, we proposed a methodology based on global similarity recognition of protein and lipid fingerprints from environmental bacteria or fungi where the spectrum/taxonomy match of the isolate is confirmed by sequencing of DNA loci (16S for bacteria and ITS for fungi) correlative to the identity. We validated our *in silico* analysis methodology by using open-source data, made available by the Robert Koch Institute (RKI, 13 orders, 38 genera, 568 fingerprints), through an R script and MetGem software [25]. Finally, we created a dataset to dereplicate bacterial (9 orders, 49 genera, 138 isolates) and fungal (21 orders, 51 genera, 230 isolates) environmental isolates, from the Bank of Natural Substances & Biodiversity (BNSB, CNRS-ICSN, Gif-sur-Yvette, France, <https://icsn.cnrs.fr/en/platforms/strain-library> (accessed on 1 January 2021) by analyzing their protein and lipid fingerprint by MALDI-ToF MS.

2. Results

2.1. Proof of Concept: t-SNE Algorithm Clusters Bacterial Protein Fingerprints in a Taxonomy-Consistent Manner

In order to evaluate the functionality of our methodology, we first used a part of the RKI dataset with the aim of comparing the global similarity of bacterial protein fingerprints. The objective was to know if the t-SNE algorithm used by the MetGem software was capable of clustering MALDI-ToF mass spectra, corresponding to protein fingerprints of microorganisms (Figure 1), in order to classify them according to their taxonomy. The t-SNE algorithm is a statistical method for visualizing high-dimensional data that captures local similarities while attempting to preserve the integrity of global structures. For that, we hijacked some MetGem functionalities in order to use the t-SNE algorithm on protein fingerprints from the RKI database. In fact, MetGem is a software able to build molecular networks and is a powerful dereplication tool in the field of metabolomics [25]. Here, MS¹ data are visualized following a pre-processing of data under a home-made script in R (for more details, see Materials and Methods, Section 4.6 and Figure S1 in Supplementary Materials). In this dataset, we have performed t-SNE visualization of 568 fingerprints belonging to 166 different bacterial strains. Each protein fingerprint of an identified clinical isolate was represented by a node (Figure 2).

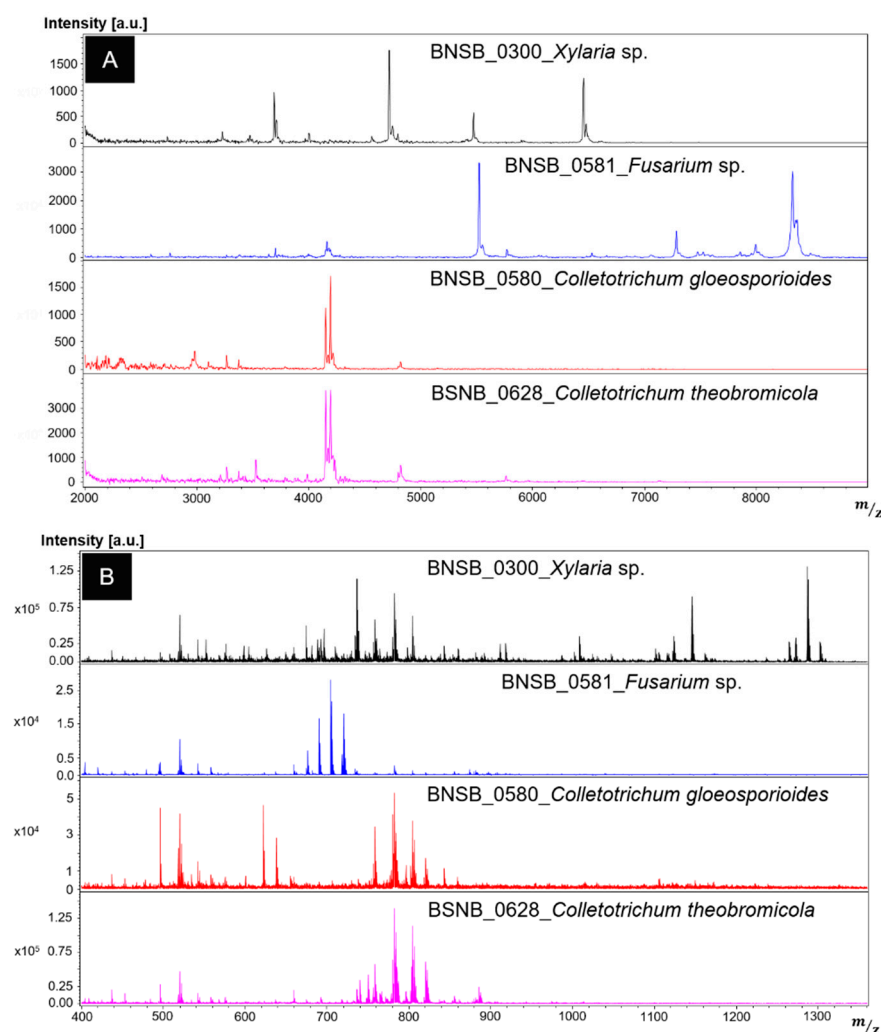


Figure 1. Examples of fungal protein fingerprints (A) and fungal lipid fingerprints (B) from BNSB.

This method led to the data aggregation into twelve clusters (Figure 2). First clusters, named A and B, are composed of Enterobacteriales in which are present the genera *Proteus*, *Serratia*, *Yersinia* (61 fingerprints—49.2% of Enterobacteriales), and *Citrobacter*, *Enterobacter*, *Escherichia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, and *Shigella* (46 fingerprints—37.1% of Enterobacteriales), respectively. Genera not present in clusters A and B are *Edwardsiella*, *Enterococcus*, and *Vibrio* which are dispersed to the north of the network or are isolated nodes (15 fingerprints—12.1% of Enterobacteriales). Cluster C (Figure 2) is only composed of the genus *Brucella* belonging to the order Rhizobiales (28 fingerprints—87.5% of Rhizobiales). The last genus of this order, i.e., *Ochrobactrum*, is located south of cluster C (4 fingerprints—12.5%). Then, cluster D (in pink, Figure 2) includes the majority of Thiotrichales, of which only the genus *Francisella* is represented (20 fingerprints—83.3% of Thiotrichales). Fingerprints corresponding to the species *Francisella guangzhouensis* are at the opposite of cluster D (4 fingerprints—16.7%). Afterwards, clusters E, F, G, H, I, J (Figure 2) are composed of Burkholderiales. Among them, the genus *Burkholderia* is present only in clusters F, G, H, I, and in their periphery (134 fingerprints—80.7% of Burkholderiales). Clusters E and J includes species of the genera *Pandora* and *Ralstonia*, respectively (24 fingerprints—14.5%). The last nodes with fingerprints from Burkholderiales are scattered in the network; east of its center (genus *Achromobacter*) or close to the J cluster (genera *Oligella* and *Xenophilus*) (8 fingerprints—4.8% of Burkholderiales). Finally, clusters K and L (Figure 2) are solely composed of fingerprints from isolates belonging to the genus *Bacillus* (91 fingerprints—53.8% of Bacillales). The last fingerprints from Bacillales are distributed in the center of the

network or in the south region and belong to the genera *Bacillus*, *Lysinibacillus*, *Paenibacillus*, and *Staphylococcus* (78 fingerprints—46.2% of Bacillales).

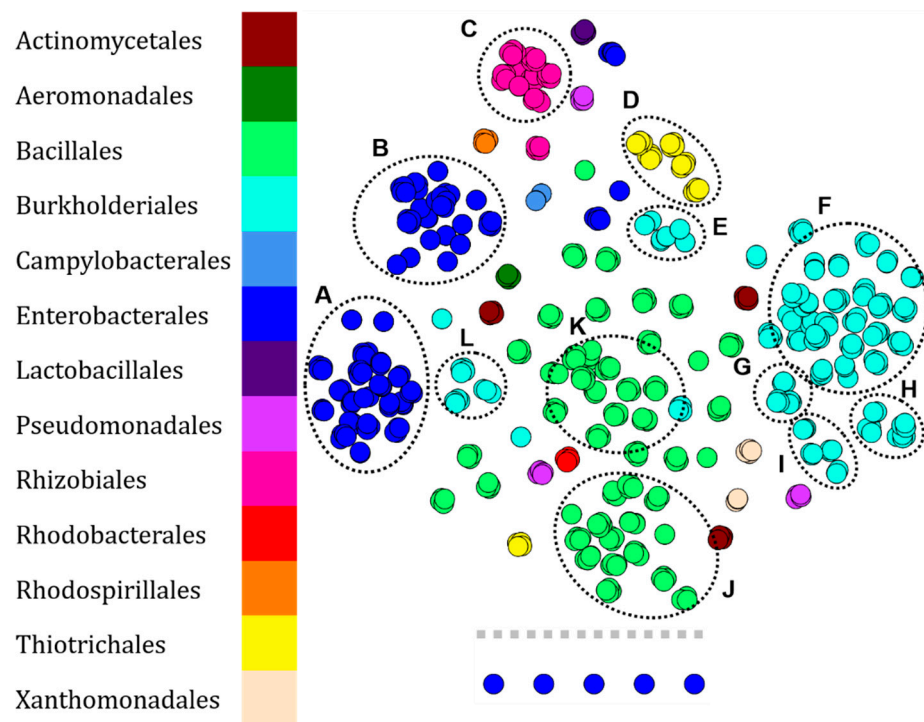


Figure 2. Visualization of 568 protein fingerprints of pathogenic bacteria (166 strains) from a part of the RKI data set by t-SNE. Each node represents the protein fingerprint of a clinical isolate identified at RKI and is colored according to the taxonomic order of the organism. The observed clusters were noted from A to L.

The protein fingerprints belonging to the rare orders (55 fingerprints), i.e., Actinomycetales, Aeromonadales, Campylobacteriales, Lactobacillales, Pseudomonadales, Rhodobacterales, Rhodospirillales, Xanthomonadales, of this dataset are scattered in the network (Figure 2) but all replicates of an isolate are close in space confirming a high robustness of the sample preparation and MALDI analysis.

In order to verify the robustness of chemotaxonomic identifications by fingerprint comparison, we added the protein fingerprints of our environmental bacterial strains to this first network.

2.2. Robustness of Chemotaxonomic Resolution by Adding Fingerprints of Environmental Isolates to the RKI Dataset

This new analysis was also obtained by visualizing the protein fingerprints of the RKI pathogenic bacteria and those of environmental bacteria through t-SNE algorithm (Figure 2). For this purpose, the initial dataset was supplemented with 138 fingerprints of bacteria whose taxonomic affiliation was previously determined by comparing the DNA sequences, encoding 16S rRNA, with those of NCBI. No new order was added to the dataset, majority orders are the same as for the RKI, i.e., Bacillales, Burkholderiales, and Enterobacteriales.

Of these 138 fingerprints, 81 nodes (58.7%) of our dataset join the RKI dataset or form a new Enterobacteriales cluster (Cluster M in Figure 3). Thus, 44 nodes are isolated (31.9%) and 13 (9.4%) are found in clusters consisting of protein fingerprints belonging to bacteria of another taxonomic order. In addition, 3 fingerprints of the genus *Vibrio* initially isolated in Figure 2 are clustered with Burkholderiales and Enterobacteriales fingerprints from the BSNB, east of the network.

In order to refine our identification by MALDI-ToF fingerprinting, we have chosen to look in parallel at the lipid fingerprints of our different strains.

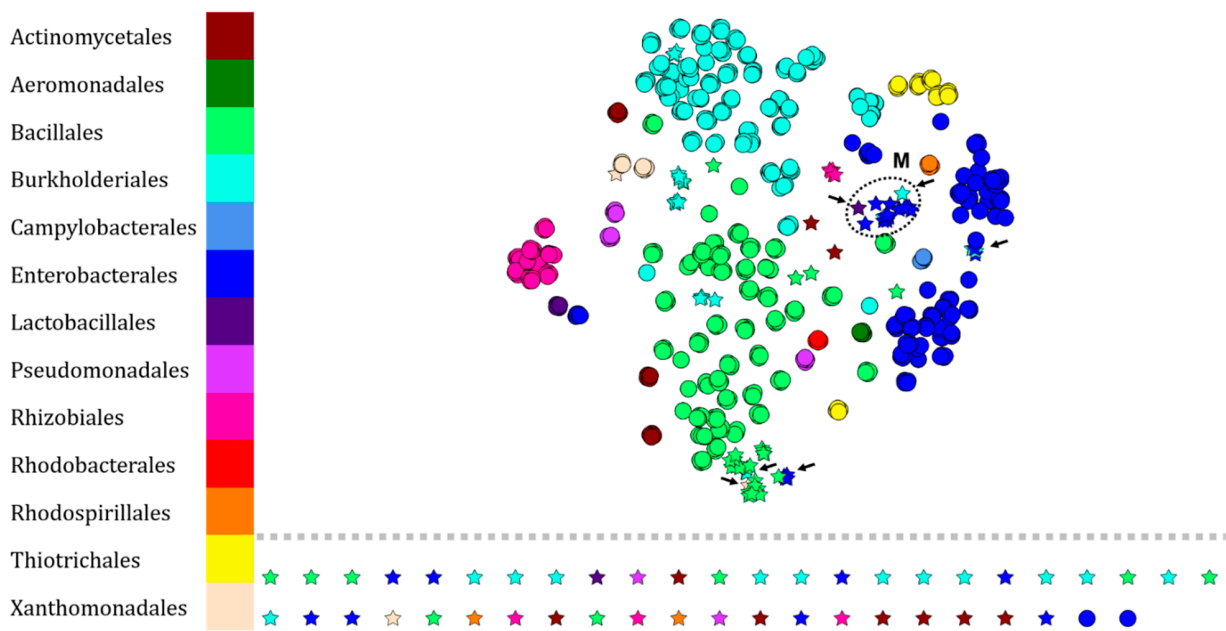


Figure 3. Visualization of 706 protein fingerprints of pathogenic bacteria (166 strains—RKI’s database; ●) and environmental bacteria (138 isolates—BNSB’s database; ★) set by t-SNE. Each node represents the protein fingerprint of an isolate and is colored according to the taxonomic order of the organism. Arrows indicate illogical attributions within clusters. The annotated cluster M consists mainly of enterobacteria.

2.3. Differentiation of Environmental Bacteria by MALDI-ToF MS Lipid Fingerprint Analysis

The purpose of this analysis was to evaluate whether lipid fingerprinting analysis of environmental microorganisms provided better differentiation than the conventional protein fingerprinting approach (Figure 4). Thus, 138 lipid fingerprints were performed on the BNSB microorganisms including the same bacteria as those for which the protein fingerprint dataset had been performed (Figure 3).

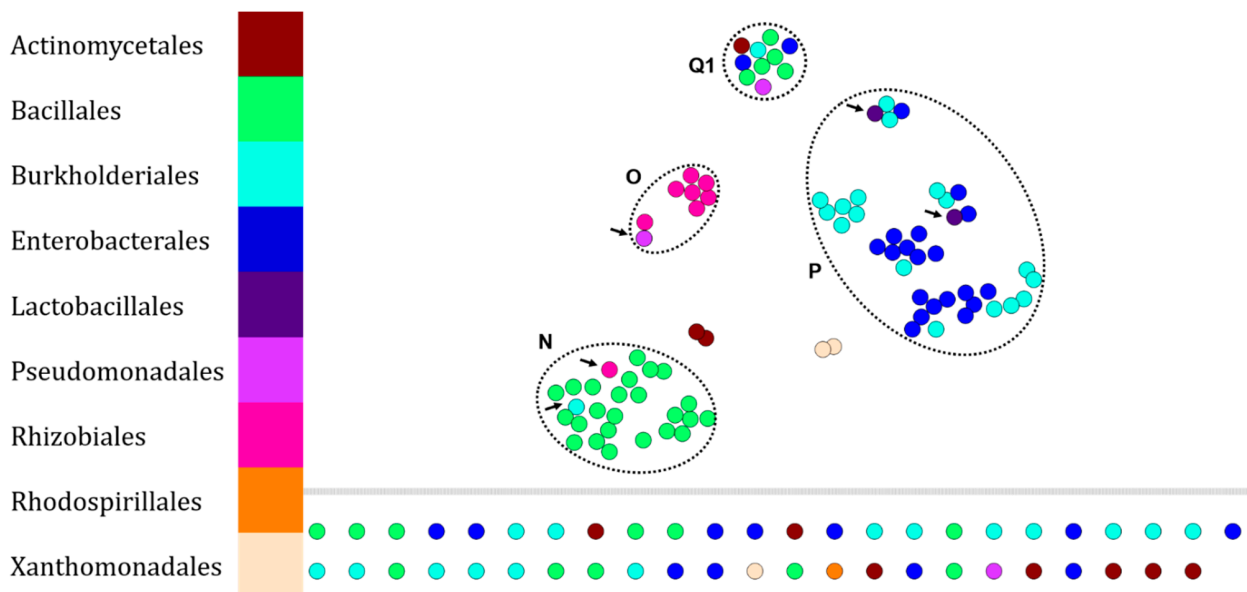


Figure 4. Visualization of 138 lipid fingerprints of environmental bacteria (138 isolates—BNSB’s database) set by t-SNE. Each node represents the lipid fingerprint of an isolate and is colored according to the taxonomic order of the organism. Arrows indicate illogical attributions within clusters (noted Q1, N, O and P).

Our chemo-informatics pipeline led to the clusterization of the major part of the dataset. Cluster N is composed mainly of Bacillales (24 fingerprints—65% of Bacillales), then Burkholderiales (1 fingerprint—2.8% of Burkholderiales) and Rhizobiales (1 fingerprint—11.1% of Rhizobiales). Cluster O is mainly composed of Rhizobiales (8 fingerprints—88.9%), then Pseudomonadales (1 fingerprint—50% of Pseudomonadales). Cluster P is composed of two main orders, i.e., Enterobacterales (18 fingerprints—56.3% of Enterobacterales) and Burkholderiales (17 fingerprints—47.2% of Burkholderiales), then Lactobacillales (2 fingerprints—100% of Lactobacillales). Cluster Q1 contains PEGs inducing clustering of lipid fingerprints from different isolates. The set of isolated nodes contains 7 Actinomycetales, 11 Bacillales, 15 Burkholderiales, 11 Enterobacterales, 1 Pseudomonadales, 1 Rhodospirillales, and 1 Xanthomonadales.

As this methodology was initially applied to our environmental bacteria of our strains collection, we wanted to evaluate if data from fungi could also be efficiently classified.

2.4. Differentiation of Environmental Fungi by MALDI-ToF MS Protein Fingerprint Analysis

Previously, we demonstrated the robustness of our methodology by analyzing protein or lipid fingerprints of environmental bacteria. Here, we applied the same methodology to protein fingerprints of environmental fungi. To this end, we first compared the protein fingerprints of 230 fungi from our collection. The resulting chemotaxonomic network is presented in Figure 5. In this network, several clusters have been identified.

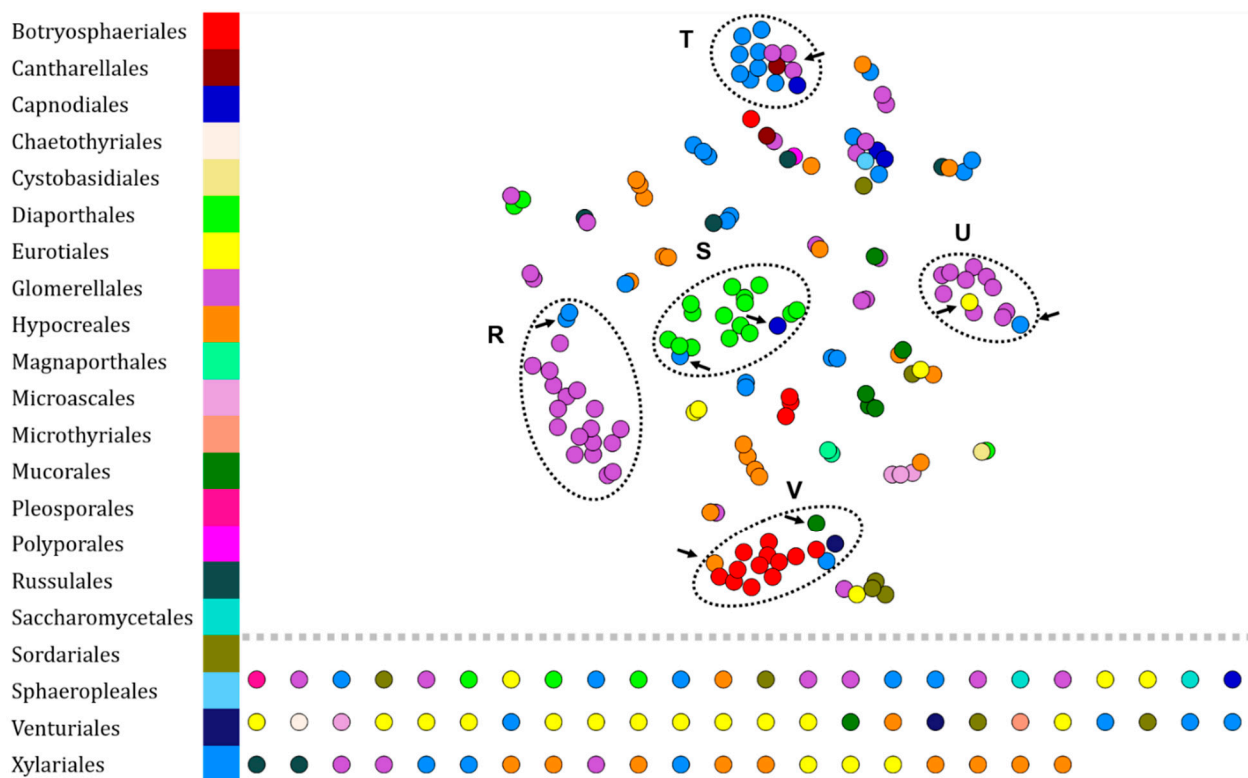


Figure 5. Visualization of 230 protein fingerprints of environmental fungi (230 isolates—BNSB’s database) set by t-SNE. Each node represents the protein fingerprint of an isolate and is colored according to the taxonomic order of the organism. Arrows indicate illogical attributions within clusters (annotated from R to V).

The cluster R is composed of 18 Glomerellales (32.7% of Glomerellales) of which 16 are *Colletotrichum gloeosporioides* (52.9%) against 1 *C. theobromicola* and 1 *C. siamense*. Finally, this cluster also contains 2 Xylariales (5% of Xylariales) protein fingerprints. Cluster S is mainly composed of Diaporthales (15 fingerprints—71.4% of Diaporthales), then of Xylariales (1 fingerprint—2.5% of Xylariales) and Capnodiales (1 fingerprint—20% of Capnodiales).

Cluster T is composed of 10 Xylariales (25% of Xylariales), 4 Glomerellales (0.7% of Glomerellales), and 1 Cantharellales (50% of Cantharellales). Group U is mainly composed of Glomerellales (10 fingerprints—18.2% of Glomerellales) of which the majority species is *Colletotrichum theobromicola* (66.7%). It also contains 1 Eurotiales (4.3%) and 1 Xylariales (2.5%). Finally, cluster V is mainly composed of Botryosphaerales (12 fingerprints—75%), then of Hypocreales (1 fingerprint—3.3%) and Xylariales (1 fingerprint—2.5%). Finally, the set of isolated nodes contains 1 Capnodiales, 1 Chaetothyriales, 3 Diaporthales, 18 Eurotiales, 9 Glomerellales, 11 Hypocreales, 1 Microascales, 1 Microthyriales, 1 Mucorales, 1 Pleosporales, 2 Russulales, 2 Saccharomycetales, 4 Sordariales, 1 Venturiales, and 12 Xylariales.

Lipid fingerprinting of the same fungal strains was also performed.

2.5. Differentiation of Environmental Fungi by MALDI-ToF MS Lipid Fingerprint Analysis

The protein fingerprint analyses performed resulted in the network shown in Figure 6. Several clusters that appeared to be specific to some orders have been identified.

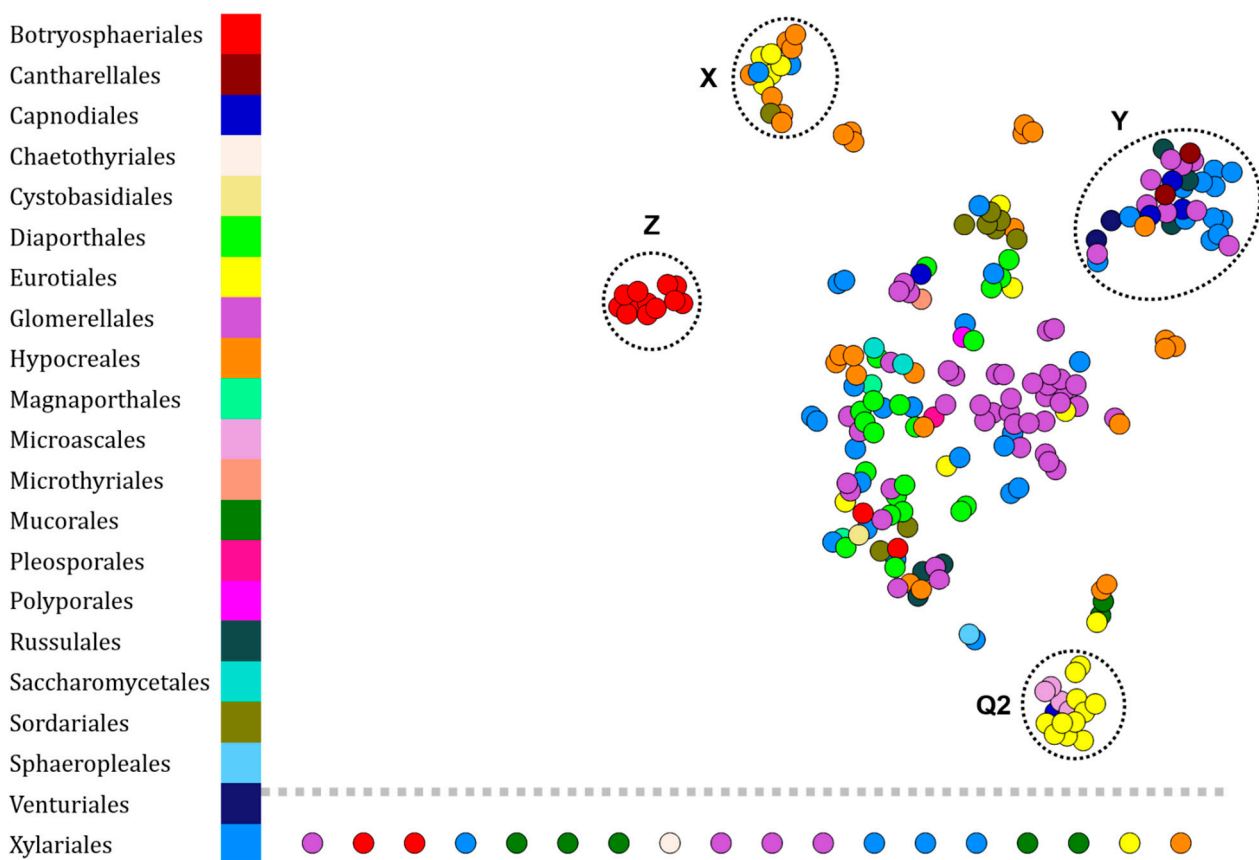


Figure 6. Visualization of 230 lipid fingerprints of environmental fungi (230 isolates—BNSB's database) set by t-SNE. Each node represents the lipid fingerprint of an isolate and is colored according to the taxonomic order of the organism. Arrows indicate illogical attributions within clusters (annotated Q2, X, Y and Z).

First, cluster X is mainly composed of Eurotiales (5 fingerprints—21.7% of Eurotiales) and Hypocreales (10 fingerprints—33.3% of Hypocreales), then of Sordariales (1 fingerprint—11.1% of Sordariales) and Xylariales (2 fingerprints—5% of Xylariales). Then, cluster Y is composed of Cantharellales (2 fingerprints—100% of Cantharellales), Capnodiales (3 fingerprints—60% of Capnodiales), Hypocreales (1 fingerprint—3.33% of Hypocreales), Russulales (3 fingerprints—50% of Russulales), Venturiales (1 fingerprint—50% of Venturiales), and Xylariales (12 fingerprints—30% of Xylariales). Cluster Z is only composed of Botryosphaerales (12 fingerprints—75% of Botryosphaerales) and all fingerprints of isolates belonging to this cluster are *Endomelanconiopsis endophytica*. The

fingerprints of the other representatives of the order Botryosphaerales are *Guignardia mangiferae* scattered in the central group or are isolated nodes. Finally, cluster Q2 contains PEGs inducing clustering of lipid fingerprints from different isolates.

The set of isolated nodes contains 2 Botryosphaerales, 1 Chaetothyriales, 1 Eurotiales, 4 Glomerellales, 1 Hypocreales, 1 Microthyriales, 5 Mucorales, and 4 Xylariales.

With the exception of the Botryosphaerales, the different orders do not cluster specifically in the constructed network.

3. Discussion

The identification and dereplication of environmental isolates is a complex objective due to the analysis of biological objects, i.e., DNA, protein, lipid, reflecting the taxonomic affiliation of the organisms studied. Today, this methodology is mostly conducted by the dereplication of protein fingerprints obtained by MALDI-ToF mass spectrometry. While this methodology is widely considered to be robust in the clinical context, it is difficult to access the identity of isolates from environmental samples due to the lack of fingerprints of organisms not pathogenic to humans in current databases.

In this study, a standardized workflow for extracting lipids and proteins from environmental microorganisms at medium throughput in a single tube of reaction medium was proposed. Then, the pre-processing of the fingerprints from the MALDI-ToF MS analyses using a “home-made” R script and the analysis of the resulting data using the MetGem software was done. Finally, the reliability of the taxonomic affiliation of the organisms was ensured by the sequencing of the loci corresponding to the 16S rDNA for bacteria or the ITS for fungi.

In a first step, we validated our data analysis method by retrieving a portion of the protein fingerprints of highly pathogenic microorganisms from the free RKI dataset (Figure 2). The t-SNE algorithm is able to group the RKI protein fingerprints according to the taxonomic affiliation of the studied organisms based on the order for Enterobacterales according to 2 clusters (A and B in Figure 2), Burkholderiales according to 6 clusters (E, F, G, H, I & J in Figure 2) and Bacillales according to 2 clusters (K and L in Figure 2). If the taxonomic resolution of this representation is validated by the spatial proximity of the nodes corresponding to the same order or genus, the results obtained are biased by the diversity and representativeness of the protein fingerprints present in the analyses. This problem is often found in environmental sampling campaigns, due to the environmental matrices used, the culture conditions of the organisms and therefore, their physiology in laboratory, without forgetting the sample preparation methodology [12,13,19,29,30].

As a second intention, we wanted to know if some protein fingerprints could be implemented and correlated with those already existing in the RKI dataset. Therefore, we acquired protein fingerprints of environmental bacteria maintained at ICSN to adjoin those of the RKI (Figure 3). Despite affiliation errors (9.4%) and lack of matches (31.9%), 58.7% of the acquired protein fingerprints were grouped in their corresponding taxonomic order or genus. Once again, the most represented orders were: Bacillales (29.8%), Burkholderiales (29.2%), and Enterobacterales (21.8%). This demonstrates the possibility of building bacterial protein fingerprint annotation networks by implementing data from the scientific community in a free manner using our workflow. In contrast, the taxonomic resolution of fungal protein fingerprints in our analysis is less efficient than that performed on our bacterial dataset. Despite a greater diversity and a larger dataset (21 orders including 51 fungal genera against 9 orders including 24 bacterial genera), few orders are annotatable except for the Glomerellales which are distributed, mostly, in 2 clusters, R corresponding to the species *C. gloeosporioides* and U to *C. theobromicola*, and the Botryosphaerales constituting the specific cluster V. Moreover, Diaporthales can be annotated according to our approach because of their majority distribution (71.4% in cluster S, Figure 4).

Today, biotyping is widely used for the dereplication of protein fingerprints, but until now, few studies have focused on the same methodology for lipid fingerprints, which are mainly used to identify fungi [30–32]. Here, we analyzed the lipid fingerprints of the

same bacteria as those used in the previous analysis (Figure 3). Again, it is possible to annotate the identity of the Bacillales (cluster N in Figure 4). However, additional data should be acquired to verify this statement on the representatives belonging to the orders Burkholderiales and Enterobacterales because their lipid fingerprints are grouped in the same cluster (cluster P, Figure 4). On the other hand, the analysis of lipid fingerprints of fungi shows a low taxonomic resolution, due to the formation of only one specific cluster of Botryosphaeriales (cluster Z in Figure 6). However, Stübiger et al. had also observed this phenomenon on a dataset of lower diversity (genera *Aspergillus*, *Penicillium*, *Saccharomyces*, and *Trichoderma*) where commercial databases (Bruker (Bremen, Germany) and Biomérieux (Marcy-l'Étoile, France)) did not satisfy the concordances between lipid fingerprints and taxonomic affiliations [32].

4. Materials and Methods

4.1. Microorganisms and Cultures

The majority of studied microorganisms are from French Guiana and were studied in previous research projects to explore their metabolome. Briefly, our environmental strain collection includes endophytes, insect-associated microorganisms and entomo- or phytopathogens [33–36]. All microorganisms were isolated on Potato Dextrose Agar (PDA, Condalab, Madrid, Spain) medium at 28 °C and stored at –80 °C in a solution of water and glycerol (20:80). Finally, their phenotypes were recorded in the Bank of Natural Substances & Biodiversity (BNSB, ICSN-CNRS, Gif-sur-Yvette, France, <https://icsn.cnrs.fr/en/platforms/strain-library> (accessed on 1 January 2021)). The microorganisms were collected with the following ABS authorizations: ABSCH-IRCC-EN-248781-1; ABSCH-IRCC-EN-248782-1 and ABSCH-IRCC-EN-245916-1 or before 2010.

4.2. Identification of Isolates

The identification process of the bacteria is performed by amplification of a portion of the gene coding for 16S rRNA (27-F: 5'-AGAGTTTGATCCTGGCTCAG-3' and 1492-R: 5'-GGTTACCTTGTTACGACTT-3') by PCR and then sequencing the amplicons (27-F: 5'-AGAGTTTGATCCTGGCTCAG-3' and 907-R: 5'-CCGTCAATTCCTTTGAGTTT-3'). The same applies to fungal identifications (primers used for PCR and sequencing: ITS1-F: 5'-AGGAGAAGTCGTAACAAGGT-3' and ITS4-R: 5'-TCCTCCGCTTATTGATATGC-3'). We use the BLASTn algorithm to compare the obtained sequences to those present on the NCBI site. If the local alignments of the sequences obtained by sequencing were significant, then the identity of the microorganism corresponded to its closest relative.

4.3. Protein and Lipid Extractions

Unless otherwise stated, all solvents used are of analytical grade. Incubation times differ according to taxonomic affiliation. Thus, proteins and lipids are extracted as soon as colonies appear, i.e., 1 to 3 days for bacteria and 7 days for fungi. We tested 3 protein extraction protocols [31,37,38] and 2 lipid extraction protocols [32,39] before using the following (data not shown): extraction protocol is adapted from Cassagne et al. and Stübiger et al. [31,34], and consists of introducing 3 to 5 bacterial colonies or a mycelium square ($\approx 5 \text{ mm}^2$) in a sterile 2 mL tube containing 300 μL H_2O Milli-Q[®], then, 200 μL of methanol (MeOH, Sigma-Aldrich, Saint Quentin Fallavier, France) and 1000 μL of methyl *tert*-butyl ether (MTBE, Sigma-Aldrich, Saint Quentin Fallavier, France). The resulting biphasic solution is mixed thoroughly during 1 min and then allowed to settle at room temperature ($\approx 20 \text{ }^\circ\text{C}$) for 5 min or until the two distinct phases reform. Decantation is prolonged overnight at $-4 \text{ }^\circ\text{C}$. Then, cell lysates are centrifuged at 13,000 rpm during 15 min and the upper organic phase is introduced into a 10 mL glass tube in order to be dried in a rotary evaporator under reduced pressure to obtain a dried lipid sample.

After the recovery of the organic phase, 900 μL of 100% ethanol (EtOH, Sigma-Aldrich, France) is introduced in the lower aqueous phase and homogenized. After 5 min centrifugation at 13,000 rpm, supernatant is discarded and residual pellet is dried at room

temperature during 30 min. Then, the pellet is incubated 5 min in 80 to 160 μL of 70% formic acid (FA, Sigma-Aldrich, Saint Quentin Fallavier, France). Finally, an equivalent volume of 100% acetonitrile (ACN, Sigma-Aldrich, Saint Quentin Fallavier, France) is added before centrifugation (13,000 rpm, 5 min) and the supernatant, containing proteins, is conserved in a 10 mL glass tube, awaiting analysis.

Each series of extraction is carried out with an extraction control containing only PDA medium and another containing no biological material. Samples are conserved at $-20\text{ }^{\circ}\text{C}$ awaiting analysis.

4.4. MALDI-ToF Sample Preparation

Each protein sample is analyzed in duplicate for which 1 μL is deposited on a spot of polished steel plate (MTP384 polish steel target, Bruker Daltonics GmbH, Bremen, Germany), then mixed with 1 μL of matrix solution and air-dried. The matrix solution for protein analysis is prepared before each series of analyses and is composed of 20 mg of α -cyano-4-hydroxycinnamic acid (CHCA, Sigma-Aldrich, France), partially solubilized in a solution of 1 mL of ACN/ H_2O /trifluoroacetic acid (50/50/2.5, $v/v/v$).

Each dried lipid sample is solubilized in 40 to 80 μL of $\text{CHCl}_3/\text{MeOH}$ (2/1, v/v), then 1 μL of the solubilized samples is diluted in a matrix solution for lipid analysis in a 1:5 ratio. The matrix solution for lipid analysis is prepared before each series of analyses and is composed of 20 mg of 2,5-dihydroxybenzoic acid (DHB, Sigma-Aldrich, France) solubilized in 500 μL of tetrahydrofuran (THF, Sigma-Aldrich, France). Finally, 1 μL of this solution is spotted on polished steel plate and left to air-dry.

4.5. MALDI-ToF Mass Spectrometry

Protein and lipid mass spectra are acquired on a MALDI-ToF/ToF UltrafleXtrem (Bruker Daltonics GmbH, Bremen, Germany) equipped with a 337 nm pulsed nitrogen laser in the linear and reflectron mode using delayed ion extraction, in positive ion mode and by accumulating, at least twice on each replicate, 1000 single laser shots. Acquisition parameters for protein analysis are the following: linear mode, delay: 250 ns, ion source voltage 1:20 kV, ion source voltage 2:18.5 kV, and mass range: m/z 2000 to 20,000. Acquisition parameters for lipid analysis are the following: reflectron mode, delay: 140 ns, ion source voltage 1:20 kV, ion source voltage 2:18 kV, reflectron analyzer 1:21.5 kV, reflectron analyzer 2:11 kV, and mass range: m/z 400 to 2000. The mass spectrometer is externally calibrated using a mixture of proteins (insulin, cytochrome C, myoglobin, and ubiquitin I) in linear mode, and a mixture of polyethylene glycol (PEG, Sigma-Aldrich, Saint Quentin Fallavier, France) in reflectron mode.

4.6. Spectra Processing

Mass spectra from the Robert Koch Institute [40] have been sorted so that taxonomic orders or genera are redundant with the taxonomic affiliation of some bacteria studied. Rare orders were deliberately added to the dataset to control the distribution of their corresponding nodes within the constructed molecular networks. In total, 568 spectra were selected, corresponding to 13 taxonomic orders among which: 12 (2.1%) are Actinomycetales, 4 (0.7%)—Aeromonadales, 169 (29.8%)—Bacillales, 166 (29.2%)—Burkholderiales, 4 (0.7%)—Campylobacterales, 124 (21.8%)—Enterobacterales, 4 (0.7%)—Lactobacillales, 12 (2.1%)—Pseudomonadales, 34 (6%)—Rhizobiales, 4 (0.7%)—Rhodobacterales, 4 (0.7%)—Rhodospirillales, 24 (4.2%)—Thiotrichales, and 7 (1.2%)—Xanthomonadales, i.e., 38 genera (Table S1 in Supplementary Material).

The bacteria of the BNSB belong to the 9 following orders: 11 (8%)—Actinomycetales, 40 (29%)—Bacillales, 36 (26.1%)—Burkholderiales, 32 (23.2%)—Enterobacterales, 2 (1.4%)—Lactobacillales, 3 (2.2%)—Pseudomonadales, 9 (6.5%)—Rhizobiales, 2 (1.4%)—Rhodospirillales, 3 (2.2%)—Xanthomonadales, i.e., 24 genera (see Table S2 in Supplementary Material). The combined RKI and BSNB protein fingerprint dataset contains 13 orders, 49 genera, and 706 fingerprints (Figure 3).

Fungi from our collection belong to the 21 following orders: 16 (7%)—Botryosphaerales, 2 (0.9%)—Cantharellales, 5 (2.2%)—Capnodiales, 1 (0.4%)—Chaetothyriales, 1 (0.4%)—Cystobasidiales, 21 (9.1%)—Diaporthales, 23 (10%)—Eurotiales, 55 (23.9%)—Glomerellales, 30 (13%)—Hypocreales, 2 (0.9%)—Magnaporthales, 4 (1.7%)—Microascales, 1 (0.4%)—Microthyriales, 7 (3%)—Mucorales, 1 (0.4%)—Pleosporales, 1 (0.4%)—Polyporales, 6 (2.6%)—Russulales, 2 (0.9%)—Saccharomycetales, 9 (3.9%)—Sordariales, 1 (0.4%)—Sphaeropleales, 2 (0.9%)—Venturiales, and 40 (17.4%)—Xylariales, i.e., 51 genera (Table S3 in Supplementary Material).

All spectra are converted in mzXML file format with MSConvert (v. 3.0.20344), a tool from ProteoWizard software [41], then are processed with a home-made script written in R software with MALDIquant and MALDIquantForeign packages [42]. Our methodology is the following: first, mass range of the protein mass spectra is adjusted to be between 3.5–20 kDa in order to avoid the implementation of background noise in our data, then the peaks with a signal to noise ratio ≥ 6 are selected, the intensities are transformed by applying a square root function, smoothed by Savitzky–Golay method (half-window size = 15) [43], then, baseline removal is conducted by Statistics-sensitive Non-linear Iterative Peak-clipping algorithm (SNIP, 30 iterations), finally, spectra are normalized using Total Ion Current (TIC) method. The last step is to create a Mascot Generic Format (MGF) file that can be interpreted by MetGem software [25]. The code is available in Figure S1 and at https://github.com/MarceauLEVASSEURCNRS/20220119_MALDI_mgf (accessed on 19 January 2022).

5. Conclusions

Our study proposed an open-source workflow including a standardized protein and lipid extraction protocol, in a tube of reaction medium, as well as a pre-processing of the data under a home-made R script, and an analysis of the fingerprints obtained by the creation of chemotaxonomic networks under the MetGem software. Our results show that this method can be used to discriminate, at medium throughput, bacterial isolates from protein or lipid fingerprints if there are sufficient representatives of a bacterial order or genus. On the other hand, the application of this same process to tropical filamentous fungi remains to be improved because of the sole clustering of orders or genera overrepresented in our data, i.e., Botryosphaerales and Glomerellales. Moreover, the *sine qua non* condition for the identification of environmental isolates will be the construction of an appropriate spectral database. To our knowledge, this work is the first to focus on the chemotaxonomy of tropical fungi. Thus, our objective is to provide the scientific community with a free and implementable spectral database of tropical microorganisms' fingerprints with a particular attention to filamentous fungi in order to improve the dereplication processes of environmental strains used following sampling campaigns. These results are encouraging and this method can be used to discriminate two different strains isolated from the same environment and the same host (in the case of symbiotic microorganisms). This first step allows to avoid redundant studies and thus to accelerate the research in the field of natural substances chemistry. Finally, this dereplication based on the calculation of cosine scores enables analysis of large datasets.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms10040831/s1>, Figure S1: R script used: generation of mgf. and csv. files (https://github.com/MarceauLEVASSEURCNRS/20220119_MALDI_mgf (accessed on 19 January 2022)), Tables S1–S6: summary of studied bacterial strains (RKI and BSNB), fungal strains (BSNB), 16S and ITS of microorganisms and R script used and parameters used for constructing chemotaxonomic networks on MetGem.

Author Contributions: Conceptualization, M.L., T.H., D.T., and V.E.; methodology, M.L., T.H., V.G., D.T. and V.E.; software, T.H. and N.E.; validation, M.L., D.T. and V.E.; formal analysis, M.L., T.H. and V.G.; investigation, M.L., T.H., D.T. and V.E.; resources, D.T. and V.E.; data curation, M.L.; writing—original draft preparation, M.L., D.T. and V.E.; writing—review and editing, M.L., N.E., D.T. and V.E.; visualization, M.L., N.E. and D.T.; supervision, D.T. and V.E.; project administration, D.T. and

V.E.; funding acquisition, D.T. and V.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work has benefited from an “Investissement d’Avenir” grant (CEBA, reference: ANR-10-LABX-0025) managed by *Agence Nationale de la Recherche*. This project has received financial support from the CNRS through the 80|Prime program. The UltrafleXtreme MALDI ToF/ToF mass spectrometer used in this study was funded by a grant from the *Région Ile-de-France* and the CNRS-ICSN (DIM Analytics Equipements Mi-Lourds 2012).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials.

Acknowledgments: We thank ICSN management who supported this project through the recruitment of an assistant engineer and Laurent Intertaglia from Bio2mar platform at Oceanological Observatory of Banyuls for the generation of sequencing data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Hyde, K.D.; Xu, J.; Rapior, S.; Jeewon, R.; Lumyong, S.; Niego, A.G.T.; Abeywickrama, P.D.; Aluthmuhandiram, J.V.S.; Brahamanage, R.S.; Brooks, S.; et al. The Amazing Potential of Fungi: 50 Ways We Can Exploit Fungi Industrially. *Fungal Divers* **2019**, *97*, 1–136. [[CrossRef](#)]
- Katz, L.; Baltz, R.H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176. [[CrossRef](#)] [[PubMed](#)]
- Briški, F.; Vuković Domanovac, M. Environmental Microbiology. *Phys. Sci. Rev.* **2017**, *2*, 20160118. [[CrossRef](#)]
- Torsvik, V. Prokaryotic Diversity—Magnitude, Dynamics, and Controlling Factors. *Science* **2002**, *296*, 1064–1066. [[CrossRef](#)] [[PubMed](#)]
- The International Natural Product Sciences Taskforce; Atanasov, A.G.; Zotchev, S.B.; Dirsch, V.M.; Supuran, C.T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug. Discov.* **2021**, *20*, 200–216. [[CrossRef](#)] [[PubMed](#)]
- Maiden, M.C.J.; van Rensburg, M.J.J.; Bray, J.E.; Earle, S.G.; Ford, S.A.; Jolley, K.A.; McCarthy, N.D. MLST Revisited: The Gene-by-Gene Approach to Bacterial Genomics. *Nat. Rev. Microbiol.* **2013**, *11*, 728–736. [[CrossRef](#)]
- Sawana, A.; Adeolu, M.; Gupta, R.S. Molecular Signatures and Phylogenomic Analysis of the Genus Burkholderia: Proposal for Division of This Genus into the Emended Genus Burkholderia Containing Pathogenic Organisms and a New Genus Paraburkholderia Gen. Nov. Harboring Environmental Species. *Front. Genet.* **2014**, *5*. [[CrossRef](#)]
- Woese, C.R.; Fox, G.E.; Pechman, K.R. Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Prokaryotic Systematics. *Int. J. Syst. Evol. Microbiol.* **1977**, *27*, 44–57. [[CrossRef](#)]
- Johnson, J.S.; Spakowicz, D.J.; Hong, B.-Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis. *Nat. Commun.* **2019**, *10*, 5029. [[CrossRef](#)]
- Schoch, C.L.; Seifert, K.A.; Huhndorf, S.; Robert, V.; Spouge, J.L.; Levesque, C.A.; Chen, W.; Fungal Barcoding Consortium; Fungal Barcoding Consortium Author List; Bolchacova, E.; et al. Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6241–6246. [[CrossRef](#)]
- Croxatto, A.; Prod’hom, G.; Greub, G. Applications of MALDI-TOF Mass Spectrometry in Clinical Diagnostic Microbiology. *FEMS Microbiol. Rev.* **2012**, *36*, 380–407. [[CrossRef](#)] [[PubMed](#)]
- Jang, K.-S.; Kim, Y.H. Rapid and Robust MALDI-TOF MS Techniques for Microbial Identification: A Brief Overview of Their Diverse Applications. *J. Microbiol.* **2018**, *56*, 209–216. [[CrossRef](#)] [[PubMed](#)]
- Santos, I.C.; Hildenbrand, Z.L.; Schug, K.A. Applications of MALDI-TOF MS in Environmental Microbiology. *Analyst* **2016**, *141*, 2827–2837. [[CrossRef](#)] [[PubMed](#)]
- Schmidt, O.; Kallow, W. Differentiation of Indoor Wood Decay Fungi with MALDI-TOF Mass Spectrometry. *Holzforschung* **2005**, *59*, 374–377. [[CrossRef](#)]
- Clark, A.E.; Kaleta, E.J.; Arora, A.; Wolk, D.M. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: A Fundamental Shift in the Routine Practice of Clinical Microbiology. *Clin. Microbiol. Rev.* **2013**, *26*, 547–603. [[CrossRef](#)]
- Costa, M.S.; Clark, C.M.; Ómarsdóttir, S.; Sanchez, L.M.; Murphy, B.T. Minimizing Taxonomic and Natural Product Redundancy in Microbial Libraries Using MALDI-TOF MS and the Bioinformatics Pipeline IDBac. *J. Nat. Prod.* **2019**, *82*, 2167–2173. [[CrossRef](#)]
- Sandrin, T.R.; Goldstein, J.E.; Schumaker, S. MALDI TOF MS Profiling of Bacteria at the Strain Level: A Review. *Mass Spectrom. Rev.* **2013**, *32*, 188–217. [[CrossRef](#)]

18. Strejcek, M.; Smrhova, T.; Junkova, P.; Uhlik, O. Whole-Cell MALDI-TOF MS Versus 16S rRNA Gene Analysis for Identification and Dereplication of Recurrent Bacterial Isolates. *Front. Microbiol.* **2018**, *9*, 1294. [CrossRef]
19. Ghyselinck, J.; Van Hoorde, K.; Hoste, B.; Heylen, K.; De Vos, P. Evaluation of MALDI-TOF MS as a Tool for High-Throughput Dereplication. *J. Microbiol. Methods* **2011**, *86*, 327–336. [CrossRef]
20. Bull, A.T.; Goodfellow, M.; Slater, J.H. Biodiversity as a source of innovation in biotechnology. *Annu. Rev. Microbiol.* **1992**, *46*, 219–252. [CrossRef]
21. Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S.S.; Wohlgemuth, G.; Barupal, D.K.; Showalter, M.R.; Arita, M.; et al. Identification of Small Molecules Using Accurate Mass MS/MS Search. *Mass Spec. Rev.* **2018**, *37*, 513–532. [CrossRef] [PubMed]
22. Wolfender, J.-L.; Litaudon, M.; Touboul, D.; Queiroz, E.F. Innovative Omics-Based Approaches for Prioritisation and Targeted Isolation of Natural Products—New Strategies for Drug Discovery. *Nat. Prod. Rep.* **2019**, *36*, 855–868. [CrossRef] [PubMed]
23. Nothias, L.-F.; Boutet-Mercey, S.; Cachet, X.; De La Torre, E.; Laboureur, L.; Gallard, J.-F.; Retailleau, P.; Brunelle, A.; Dorrestein, P.C.; Costa, J.; et al. Environmentally Friendly Procedure Based on Supercritical Fluid Chromatography and Tandem Mass Spectrometry Molecular Networking for the Discovery of Potent Antiviral Compounds from *Euphorbia Semiperfoliata*. *J. Nat. Prod.* **2017**, *80*, 2620–2629. [CrossRef] [PubMed]
24. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B.S.; Yang, J.Y.; Kersten, R.D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J.M.; et al. Mass Spectral Molecular Networking of Living Microbial Colonies. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743–E1752. [CrossRef]
25. Olivon, F.; Elie, N.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. MetGem Software for the Generation of Molecular Networks Based on the T-SNE Algorithm. *Anal. Chem.* **2018**, *90*, 13900–13908. [CrossRef]
26. Elie, N.; Santerre, C.; Touboul, D. Generation of a Molecular Network from Electron Ionization Mass Spectrometry Data by Combining MZmine2 and MetGem Software. *Anal. Chem.* **2019**, *91*, 11489–11492. [CrossRef]
27. Dumolin, C.; Aerts, M.; Verheyde, B.; Schellaert, S.; Vandamme, T.; Van der Jeugt, F.; De Canck, E.; Cnockaert, M.; Wieme, A.D.; Cleenwerck, I.; et al. Introducing SPeDE: High-Throughput Dereplication and Accurate Determination of Microbial Diversity from Matrix-Assisted Laser Desorption–Ionization Time of Flight Mass Spectrometry Data. *mSystems* **2019**, *4*, e00437-19. [CrossRef]
28. Dumolin, C.; Peeters, C.; De Canck, E.; Boon, N.; Vandamme, P. Network Analysis Based on Unique Spectral Features Enables an Efficient Selection of Genomically Diverse Operational Isolation Units. *Microorganisms* **2021**, *9*, 416. [CrossRef]
29. Rahi, P.; Prakash, O.; Shouche, Y.S. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass-Spectrometry (MALDI-TOF MS) Based Microbial Identifications: Challenges and Scopes for Microbial Ecologists. *Front. Microbiol.* **2016**, *7*. [CrossRef]
30. Cassagne, C.; Normand, A.-C.; L’Ollivier, C.; Ranque, S.; Piarroux, R. Performance of MALDI-TOF MS Platforms for Fungal Identification. *Mycoses* **2016**, *59*, 678–690. [CrossRef]
31. Cassagne, C.; Ranque, S.; Normand, A.-C.; Fourquet, P.; Thiebault, S.; Planard, C.; Hendrickx, M.; Piarroux, R. Mould Routine Identification in the Clinical Laboratory by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *PLoS ONE* **2011**, *6*, e28425. [CrossRef] [PubMed]
32. Stübiger, G.; Wuczowski, M.; Mancera, L.; Lopandic, K.; Sterflinger, K.; Belgacem, O. Characterization of Yeasts and Filamentous Fungi Using MALDI Lipid Phenotyping. *J. Microbiol. Methods* **2016**, *130*, 27–37. [CrossRef] [PubMed]
33. Barthélemy, M.; Guérineau, V.; Genta-Jouve, G.; Roy, M.; Chave, J.; Guillot, R.; Pellissier, L.; Wolfender, J.-L.; Stien, D.; Eparvier, V.; et al. Identification and Dereplication of Endophytic Colletotrichum Strains by MALDI TOF Mass Spectrometry and Molecular Networking. *Sci. Rep.* **2020**, *10*, 19788. [CrossRef] [PubMed]
34. Brel, O.; Touré, S.; Levasseur, M.; Lechat, C.; Pellissier, L.; Wolfender, J.-L.; Van-Elislande, E.; Litaudon, M.; Dusfour, I.; Stien, D.; et al. Paecilostatin Derivatives as Potent Antimicrobial Agents from *Isaria Farinosa*. *J. Nat. Prod.* **2020**, *83*, 2915–2922. [CrossRef] [PubMed]
35. Hebra, T.; Elie, N.; Poyer, S.; Van Elislande, E.; Touboul, D.; Eparvier, V. Dereplication, Annotation, and Characterization of 74 Potential Antimicrobial Metabolites from *Penicillium Sclerotiorum* Using t-SNE Molecular Networks. *Metabolites* **2021**, *11*, 444. [CrossRef] [PubMed]
36. Mai, P.-Y.; Levasseur, M.; Buisson, D.; Touboul, D.; Eparvier, V. Identification of Antimicrobial Compounds from *Sandwithia Guyanensis*-Associated Endophyte Using Molecular Network Approach. *Plants* **2019**, *9*, 47. [CrossRef]
37. Becker, P.T.; de Bel, A.; Martiny, D.; Ranque, S.; Piarroux, R.; Cassagne, C.; Detandt, M.; Hendrickx, M. Identification of Filamentous Fungi Isolates by MALDI-TOF Mass Spectrometry: Clinical Evaluation of an Extended Reference Spectra Library. *Med. Mycol.* **2014**, *52*, 826–834. [CrossRef]
38. Mancini, V.; Dapporto, L.; Baracchi, D.; Luchi, N.; Turillazzi, S.; Capretti, P. Phenotypic Characterization of Cryptic *Diplodia* Species by MALDI-TOF MS and the Bias of Mycelium Age. *For. Path.* **2013**, *43*, 455–461. [CrossRef]
39. Calvano, C.D.; Zamboni, C.G.; Palmisano, F. Lipid Fingerprinting of Gram-Positive Lactobacilli by Intact Cells—Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using a Proton Sponge Based Matrix: Lipid Fingerprinting of Gram-Positive Lactobacilli by Intact Cells. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1757–1764. [CrossRef]
40. Lasch, P.; Stämmle, M.; Schneider, A. Version 3 (20181130) of the MALDI-TOF Mass Spectrometry Database for Identification and Classification of Highly Pathogenic Microorganisms from the Robert Koch-Institute (RKI) 2018. Zenodo. Available online: <https://zenodo.org/record/1880975#.Y1t20DURXIU> (accessed on 15 February 2019). [CrossRef]
41. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24*, 2534–2536. [CrossRef]

-
42. Gibb, S.; Strimmer, K. MALDIquant: A Versatile R Package for the Analysis of Mass Spectrometry Data. *Bioinformatics* **2012**, *28*, 2270–2271. [[CrossRef](#)] [[PubMed](#)]
 43. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [[CrossRef](#)]