I COBBIND
I CONGRESSO BRASILEIRO DE BIOTECNOLOGIA INDUSTRIAL

XXIV SINAFERM ++ XV SHEB ++ XV ENZITEC   FLORIANÓPOLIS ++ 2024

Creating connections between biotechnology and industrial sustainability

August 25 to 28, 2024
Costão do Santinho Resort, Florianópolis, SC, Brazil

INDUSTRIAL MICROBIOLOGY: PROSPECTING AND APPLIED MOLECULAR BIOLOGY

MINING BIOPIGMENTS IN PUBLIC DATASETS THROUGH UNTARGETED METABOLOMICS

Fernanda de Oliveira[1*], Caio A. Lima[2], Júlio G. De Lima[2], Leonardo H. Rotta[2], Valéria C. Santos-Ebinuma[2] Pieter C. Dorrestein[3], & Silvio S. da Silva[1]

[1] Bioprocesses and sustainable products laboratory. Department of Biotechnology, Engineering School of Lorena, USP, Lorena, Brazil.

[2] Department of Bioprocess Engineering and Biotechnology, School of Pharmaceutical Sciences, UNESP, Araraquara, Brazil.

[3] Skaggs School of Pharmacy & Pharmaceutical Sciences. Department of Pharmacology & Pediatrics, University of California, San Diego, USA.

* Corresponding author's email address: fernanda.oliveira1@unesp.br

ABSTRACT

Natural products discovery is progressively employing -omics technologies to guide molecular exploration. Most of the publications on biopigments belongs to the microbiology field, providing a good overview on the topic with an impressive array of molecular structures, but a detailed description of the available structures is largely missing. There are many opportunities aiming the valorization of fungi-derived pigments, including novel chemical structures discovery for the development of high-value niche markets. Given the underexplored chemistry of fungi, a strategy was described for chemotaxonomic investigation using fungi species of the Trichocomaceae family grown in different culture media as a model for novel pigment discovery. The workflow was based on MS/MS untargeted metabolomics, spectral searches in publicly available data, and in silico annotation and classification tools. The chemical structural information was represented as a network and chemical relationships were visualized. The metabolomic analyses revealed that different culture media significantly affected the chemical profiles of intra- and extracellular pigments in fungi species, influencing the chemotaxonomic results. Spectral searches within public databases revealed several genera-specific molecular families, being potential chemical markers for these taxa. The in silico annotation propagation tools brought out at least two novel azaphilone pigments, expanding the chemical space for Trichocomaceae family.

Keywords: Metabolomics. GNPS. Molecular Networking. Biopigments.

1 INTRODUCTION

The pigment industry is facing a change of perspectives and sustainable solutions for safer processes has brought attention to biopigments. Replacing chemical pigments is a hard task. Significant challenges and potential solutions that limit the implementation of bio-based pigments falls into the structural diversity of petroleum-based pigments discovered over the past century[1]. Most of the publications on biopigments belongs to the microbiology field, providing a good overview on the topic, occasionally with an impressive array of molecular structures, but a detailed description of the available structures is largely missing[2]. The vast family of azaphilone pigments constitute a significant portion of fungal secondary metabolites prevalent in various fungal species, including *Penicillium*, *Aspergillus*, *Talaromyces*, *Chaetomium*, and *Monascus*[3]. With many pigment producers remaining underexplored, numerous opportunities for future research exist, including the novel chemical structures dereplication, paving the way for the development of high-value niche markets. The Global Natural Products Social Molecular Networking (GNPS) (https://gnps.ucsd.edu/) is an online tool to perform the elucidation of small molecules and natural products based on MS^2 spectral alignment. In a pairwise MS^2 spectrum alignment, each spectrum in a given dataset is compared against every other, and a network of MS^2 spectral relations is obtained, from which molecular networks are created. GNPS provides the ability to analyze a dataset and to compare it to all publicly available data, enabling the mapping of the chemical diversity observed in an untargeted mass spectrometry experiment[4]. Propagation is done through manual inspection of MS^2 spectra connected in the spectral networks and one of the high throughput approaches used to annotate an unknown fragmentation mass spectrum is through the use of in silico predictions. We performed metabolomics analyses in combination with in silico fragmentation predictors and evaluated (i) the impact of different nitrogen in culture media for chemotaxonomic investigations; (ii) the metabolites annotated based on spectral matches and in silico tools; and (iii) how the chemical diversity evolved in fungi species of Trichocomaceae family. By using metabolomics approach for characterizing the composition of the holistic chemical structures, we were able to provide a comprehensive overview of the metabolites produced by the Trichocomaceae family.

2 MATERIAL & METHODS

The chemical diversity of two strains of fungi from *Trichocomaceae* family was assessed by varying the nitrogen source (monosodium glutamate [MSG], yeast extract [YE], meat peptone [MP], meat extract [ME], ammonium sulfate [NH4-SO], and ammonium nitrate [NH4-NO]) in a medium containing glucose, 12 mgL^{-1} of $MgSO_4$, 15 mgL^{-1} of $CaCl_2$, and 10 mgL^{-1} of $FeSO_4$. The C:N relation was kept in 6.5. To this purpose, 5 mycelial agar discs (8 mm diameter) of PDA plates were punched out and transferred to 50 mL of submerged culture medium in 250-mL Erlenmeyer flasks. The pH of the media was adjusted to 5.0 before sterilization at 121°C for 15 min. Incubation of the flasks was performed for 168 h in incubator shaker under agitation (150 rpm) at 30°C. The extracts were prepared by harvesting the biomass under filtration using a filter paper (80 g/m2; Whatman, UK). The biomass was frozen and stored in an ultra-freezer at −80°C for further intracellular pigments extraction. The supernatant was filtrated using 0.45-μm filter (Millipore) and stored in an ultra-freezer at −80°C for further extracellular pigments extraction. The previously stored biomass and supernatant were lyophilized and both were extracted using 50:50 *v/v* of LC grade methanol/ LC

grade water and homogenized in ultrasonic bath. Both supernatant and biomass extracts were centrifuged at 5,000g for 10 minutes at 4°C. An aliquot of 200 µL of sample was transferred to a 96 well deep plates and dried down using a centrifugal low pressure system (Labconco CentriVap, United States). Dried samples were sealed and stored in -80 ºC until analysis. The dried extracts were resuspended in 200 µL of resuspension buffer (50:50 *v/v* MeOH:H2O water with 1.0 µM Sulfamethazine as internal standard [to monitor sample injection during the Ultra High Performance Liquid Chromatography (UHPLC)-tandem Mass Spectrometry (MS/MS) data acquisition]. The metabolomic analyses were performed in a Vanquish UHPLC system coupled to a Q-Exactive orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), controlled by Thermo SII for Xcalibur software (Thermo Fisher Scientific, Waltham, MA, USA). Acquired raw spectral data were converted to .mzML files using (Proteowizard, MSconvert, version 4.7.2)[4]. Data preprocessing was performed using Mzmine (version 3.9.0)[5]. Thereafter, the data were uploaded to the MassIVE repository (data set MSV000094045). MS^2 fragmentation spectra from MZmine pre-processing were uploaded to GNPS to perform Feature-Based Molecular Networking[6] on the GNPS platform. All spectra contained in the molecular networks were compared to the reference spectra available in the GNPS spectral librarie[7], and a cosine of 0.7 and a minimum of 4 MS^2 matches were applied. microbeMASST[8] and other *in silico* tools (see Figure 1) were employed to make MS^2 single searches for the molecules detected in higher abundance. The molecular network file was visualized using Cytoscape 3.9.1[9], in which the nodes correspond to ion features, while the edges between the nodes represent the MS^2 cosine scores calculated. Multivariate data analysis was implemented to rapidly identify polyketide pigments across the microbial cultures. The feature matrix from MZmine pre-processing (containing the peak areas of each found metabolic feature in each sample) was performed using MetaboAnalystR 6.0[10] for MS^2 spectra processing and quantification, statistical analysis, and functional interpretation. For a better understanding of the workflow developed in this work, Figure 1 is presented, showing the different stages performed for the metabolomics analyses.
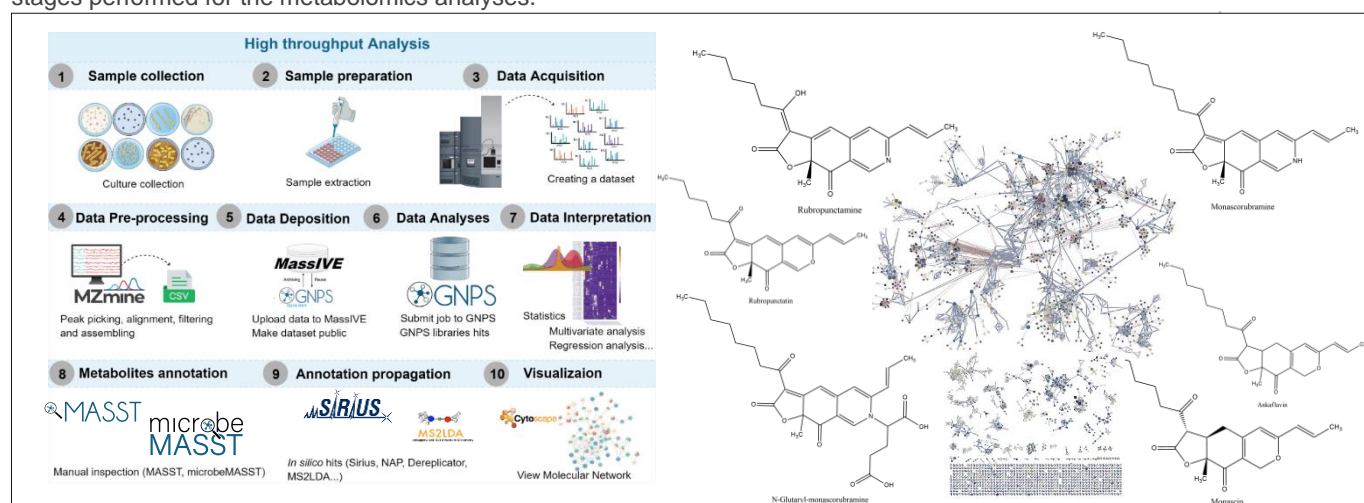
Figure 1. Workflow for the high throughput analysis performed for untargeted metabolomics used in this work and the azaphilone molecular family in the molecular network showing examples of the molecules annotated.

3 RESULTS & DISCUSSION

To observe the chemical space provided by the metabolomic profiles obtained by culturing the two fungi strains in different nitrogen sources, Principal Component Analysis (PCoA) correlating culture media (Figure 2a) and colorant location (Figure 2b) were created, a heat map correlating fungi species were also created (Figure 2c).
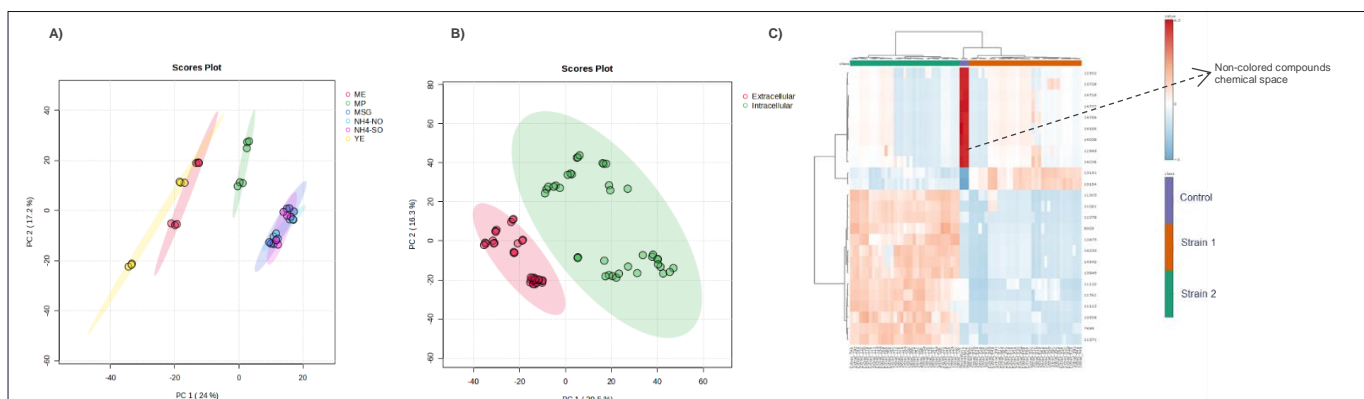
Figure 2. Three-dimensional Principal Coordinates Analysis (PCoA) plots of the samples analyzed based on A) Culture media; B) Colorant location determined by Bray–Curtis distance metric; and C) heat map for species correlation. The percentage of variance explained by the principal coordinates is presented on each axis.

PCoA was used for qualitative clustering and permutational multivariate analysis of variance (PERMANOVA) for significance testing. In summary, PCoA measures correlations among the samples by calculating the distances among them, and the way these distances are calculated can result in different clustering trends in the plots. For culture media, even though most of the features detected are shared among media composed of MSG, NH4-SO and NH4-NO, the compounds' relative abundance may significantly vary resulting in different metabolomic profiles (PERMANOVA F=18.71, p=0.001). On the other hand, the metabolomic profile obtained for extra and intracellular features did not show any clear correlation (PERMANOVA F=40.39,

p=0.001). The clustering trends observed for the species resulted in representations of how different chemical profiles can be obtained from extracts originated between related but very specific species (PERMANOVA F=28.59, p=0.001). Interestingly, the control shares more features with strain 1 than with strain 2. The control it is a well known azaphilone colorants producer and was included in this dataset just to improve annotation propagation.

Since the chemical profiles obtained for the fungi extracts showed a promising correlation within the different culture media, we made an in-depth investigation of the top intense features detected in each culture media by using *in silico* tools of fragments annotation. The most abundant nitrogen source observed included MP, ME, YE, and MSG, accounting for 59% of all read counts study-wide. Beta diversity was assessed using weighted UniFrac distance to summarize the microbial composition among the groups. The most intense features are shared among MP, ME, YE, and both ammonium sources share some intense read counts. As expected, MSG induced the most different features and the same chemical profile was detected for strain 2 and strain 1. The same observation can be done for MP, the same chemical profile was detected for both strains. From these results, we observed that the features are distributed into two main groups: organic nitrogen (MP, ME, and YE) and inorganic nitrogen sources (MSG, NH4-NO, and NH4-SO). This distribution can significantly impact the results of multivariate analysis.

Overall, the processed mass spectrometry data resulted in 9,541 MS^1 features (i.e., a detected signal with *m/z* and retention time corresponding to a detected molecule). MS^2 spectra were collected for all MS^1 features, which were represented in a molecular network (Figure 1) based on spectral similarity. Around 804 hits of the MS^2 spectra were annotated by library matches against the reference GNPS public libraries, resulting in 8.4% of the detected chemical spaces, respectively. These matches were manually evaluated and compared with the literature, resulting in level 2 or 3 annotations. In addition, the most intense features correspondent to azaphilones (after search for spectral similarity with the control annotated compounds) were also searched in the microbeMASTT workflow and other tools (data not shown). The number of nodes in the molecular families was equal 5046, showing 7170 of edges and 2967 of connected components. The number of annotated nodes was equal 114. Some examples of annotated molecules are illustrated in Figure 1. Also, the two novel molecules detected in strain 2 were not found in other public datasets. Networks containing library matches corresponding to azaphilones were prioritized and the chemical structures manually annotated are shown in Table 1.

Table 1. Chemical structures annotated d based on spectral matches within the GNPS platform.

Compound	Mass (g/mol)	Canonical Smiles
Rubropunctamine	353.4	CCCCC/C(=C\1/C2=CC3=CC(=NC=C3C(=O)[C@@]2(OC1=O)C)/C=C/C)/O
Rubropunctatin	354.4	CCCCCC(=O)C1=C2C=C3C=C(OC=C3C(=O)[C@@]2(OC1=O)C)/C=C/C
Monasfluore A	356.42	CCCCCC(=O)C1C2C3=COC(=CC3=CC(=O)C2(OC1=O)C)C=CC
Monascin	358.4	CCCCCC(=O)[C@@H]1[C@H]2CC3=C(COC(=C3)/C=C/C)C(=O)[C@@]2(OC1=O)C
Monasfluol A	360.406	CCCCCC(=O)C1C2C(OC1=O)C(=O)C=C1C=C(CC(C)O)OC=C21
Monascorubramine	381.46	C/C=C/C(/NC=C1C2=O)=CC1=CC([C@]23C)=C(C(O3)=O)C(CCCCCCC)=O
Monascorubrin	382.4	CCCCCCCC(=O)C1=C2C=C3C=C(OC=C3C(=O)[C@@]2(OC1=O)C)/C=C/C
Monasfluore B	384.47	CCCCCCC(=O)C1C2C3=COC(=CC3=CC(=O)C2(OC1=O)C)C=CC
Ankaflavin	386.5	CCCCCCC(=O)C1C2C3=C(COC(=C3)C=CC)C(=O)C2(OC1=O)C
Monasfluol B	388.46	CCCCCCC(=O)C1C2C(C(=O)C=C3C2=COC(=C3)CC(C)O)OC1=O
PP-V	411.17	O=C(O1)C(C(CCCCCCC)=O)=C2C1(C)C(C3=C[NH2]C(/C=C/C([O-])=O)=CC3=C2)=O
PP-O	412.15	O=C(O1)C(C(CCCCCCC)=O)=C2C1(C)C3=COC(/C=C/C(O)=O)=CC3=C2)=O
Glycylrubropunctatin	413.5	CCCCCC(=O)C1=C2CC3=C(CN(C(=C3)C=CC)CC(=O)O)C(=O)C2(OC1=O)C
Sequoiamonascin A	418.5	CCCCCCCC[C@@H]1[C@H]2CC3=C(CO[C@@]4(C3)CC(=O)O4)C(=O)O4)C(=O)[C@@]2(OC1=O)C)O
Sequoiamonascin B	418.5	CCCCCCCC([C@@H]1[C@H]2CC3=C(CO[C@]4(C3)C=CC(=O)O4)C(=O)[C@@]2(OC1=O)C)O
Sequoiamonascin C	418.6	CCCCCCCC([C@@H]1[C@H]2CC3=C(COC(=C3)/C=C/C(=O)O)C(=O)[C@@]2(OC1=O)C)O
Sequoiamonascin D	471.55	CCCCCC[C@@H](C)C(=O)C1=C2C(=C3C=C(N(C=C3C(=O)[C@@]2(OC1=O)C)CCO)C)C(=O)OC
PP-R	425.22	O=C(O1)C(C(CCCCCCC)=O)=C2C1(C)C(C3=CN(CCO)C(/C=C/C)=CC3=C2)=O
N-glutamyl monascorubramine	511.6	CCCCCCCC(=O)C1=C2C=C3C=C(N(C=C3C(=O)C2(OC1=O)C)C(CCC(=O)O)C(=O)O)C=CC
N-Glutaryl-monascorubramine	511.2	CCCCCCCC(=O)C1=C2C=C3C=C(N(C=C3C(=O)C2(OC1=O)C)CCC(=O)O)C(=O)O)C=CC
Z-hydroxyhexanoic acid-amestolkin*	541.273	O=C(C1=C[NH+](C(CCO)CCC(O)=O)C(/C=C\C(O)=O)=CC1=C2)C3(C)C2=C(C(CCCCCCC)=O)C(O3)=O
E-hydroxyhexanoic acid-amestolkin*	541.273	O=C(C1=C[NH+](C(CCO)CCC(O)=O)C(/C=C/C(O)=O)=CC1=C2)C3(C)C2=C(C(CCCCCCC)=O)C(O3)=O

*Novel compounds

4 CONCLUSION

Overall, the metabolomics study provided a valuable insight into azaphilone colored metabolic profile in Trichocomaceae family. It also identified new targets for colored azaphilones and expanded the chemical space of colorants, which may be useful for developing biotechnology industries. Approximately 23 chemical structures from the class of azaphilone pigments were annotated against MS^2 databases, being two of them novel colored azaphilones and the annotation of novel compounds are still in progress. Furthermore, relative quantification of precursor ions was conducted to correlate strains, providing statistical interpretation of the most abundant features identified within the complex mixture of compounds. However, due to the large number of hits found, dereplication and prioritization of the hits found after conducting a high-throughput screening it is crucial to expand the biologically relevant chemical space of azaphilone fungi colorants.

REFERENCES

[1] NUMAN, M., BASHIR, S., MUMTAZ, R., et al. 2018. Biotech. 8:207.
[2] SEN, T., BARROW, C. J., DESHMUKH, S. K. 2019. Front. Nutr. 6 (1). 1–14.
[3] DE OLIVEIRA, F., FERREIRA, L. C., NETO, et al. 2020. Biochem. Eng. J. 161. 107694.
[4] CHAMBERS, M. C., MACLEAN, B., et al. 2012. Nat Biotechnol. doi: 10.1038/nbt.2377.
[5] SCHMID, R., HEUCKEROTH, S., KORF, A., et al. 2023. doi: 10.1038/s41587-023-01690-2.
[6] NOTHIAS, L-F., PETRAS, D., SCHMID, R., et al. 2020. Nat Methods. doi: 10.1038/s41592-020-0933-6.
[7] WANG, M., CARVER, J. J., PHELAN, V. V., et al. 2016. doi: 10.1038/nbt.3597.
[8] DORRESTEIN, P., ZUFFA, S., SCHMID, R., et al. 2023. bioRxiv. doi: 10.21203/rs.3.rs-3189768/v1.
[9] SHANNON, P., MARKIEL, A., OZIER, O., et al. 2003. Genome Res. doi: 10.1101/gr.1239303.
[10] PANG, Z., LU, Y., ZHOU, G., HUI, F., et al. 2024. Nucleic Acids Research. doi: 10.1093/nar/gkae253.

ACKNOWLEDGEMENTS