Discrimination Between *Cryptococcus neoformans* varieties

Manuel J. Arroyo[1], Margarita E. Zvezdanova[2], Belén Rodríguez-Sánchez[2], Luis Mancera[1]

Abstract

Matrix-Assisted Laser Desorption Ionization - Time of Flight Mass Spectrometry (MALDI-TOF MS) has emerged as a promising technology for the rapid and reliable identification of yeasts. In this work, we develop a rapid and reliable tool for the correct discrimination of *Cryptococcus neoformans* var. *grubii* and var. neoformans. We also show good discrimination of hybrids variety samples. Differentiation among these varieties is important to define the epidemiology of the infection. Clover MS Data Analysis software is capable to help and find reference peaks of these microorganism to serve as potential biomarkers. In this document we show how the platform can correctly classify these varieties from a previously created prediction base.

Introduction

Cryptococcus neoformans is one of the causative agents of cryptococcosis disease. It has been associated with meningitis in immunosuppressed patients as well. Subspecies identification is important in order to establish the epidemiology, virulence and susceptibility pattern to the commonly used antifungal drugs. Molecular techniques have shown to be accurate and robust, although the whole procedure is cumbersome, time consuming, and delays the final identification. On the other hand, the high number of different molecular methods and the lack of a consensus technique for their identification has led to the evaluation of alternative tools that solve these disadvantages.

Materials and methods

For the construction of a prediction base in this study, we used n=69 samples from the Hospital Gregorio-Marañón, in Madrid, Spain (34 var. *grubii*, 28 var. *hybrid* and 7 var. *neoformans*). The sample processing method applied consisted of a mechanical disruption step followed by a standard protein extraction.

Each sample was obtained from three replicated spectra which were used to make an average spectrum of each sample. These main spectrums were preprocessed in the platform in order to reduce the baseline and noise to make a clean peak matrix. The optimal parameters of baseline subtraction and smoothing depend on the quality of the spectrum obtained. Table 1 shows the optimized parameters used for these data.

Baseline Subtraction	Type: Top-Hat filter, factor: 0.02
Noise Reduction	Type: Saviztky-Golay, window length: 25, polynomial order:3

Table 1. Optimized pre-processing parameters

After pre-processing, we detected peaks by mass position using the following reference mass list (all units in m/z): 2488, 2870, 3114, 3243, 3363, 4059, 4237, 4821, 4852, 5452, 5551, 5704, 6486, 6501, 6577, 6591, 6727, 6798 and 7100. These peaks were chosen visually and their reproducibility studied across all samples.

Peak Finding	Type: By mass position, constant Mass tolerance: 5Da, linear Mass tolerance: 0ppm, prominence: 0, width: 0

Table 2. Parameter for peak detection

Experimental

In the case of *Cryptococcus*, a two-step discrimination method has been developed making a two different peak matrix.

In the first step, a peak matrix is generated using all the spectra which were preprocessed as shown in table 1 and aligned with a 5 Da tolerance, and then normalized by the intensity at mass 2488m/z. This peak matrix is used as entry for a hierarchical clustering analysis, as shown in Figure 1. This

[1] Clover Bioanalytical Software, S.L. Granada, Spain

[2] Clinical Microbiology and Infectious Diseases Dep., Hospital General Universitario Gregorio Marañón, Madrid, Spain

shows a two well defined cluster and help knowing if the parametrers used in the peak matrix generation were a good choice.

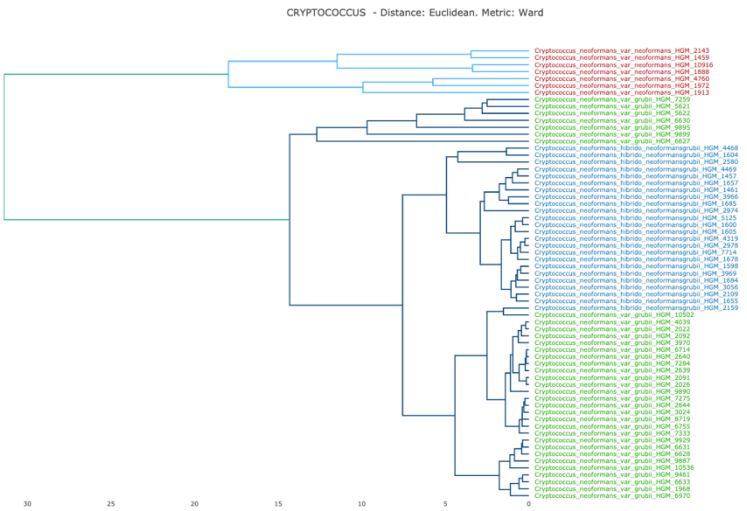

Figure 1. Dendrogram as a result of the first step peak matrix when it is analyzed by a Hierarchical clustering module. Two big groups are shown, var. *neoformans* and var. *grubii/hybrid*.

In this step a dendrogram was obtained showing two different clusters where *Cryptococcus* var. *neoformans* samples were well distinguished from *Cryptococcus* var. *grubii* and hybrids.

In the second step, a new peak matrix was built to achieve a better separation of *Cryptococcus* var. *grubii* from the *hybrids*. This peak matrix was generated without *Cryptococcus* var. *neoformans* samples (already separated and identified before). The same method and parameters as for the previous classification was used. Nevertheless, in the "Normalization" step of the peak matrix generation process, the normalization applied are Total Ion Current (TIC). For this new peak matrix, we can evaluate the choice of parameters by analyzing with PCA and hierarchical clustering algorithms. This shows that the two remaining groups are separated correctly (Figure 2).

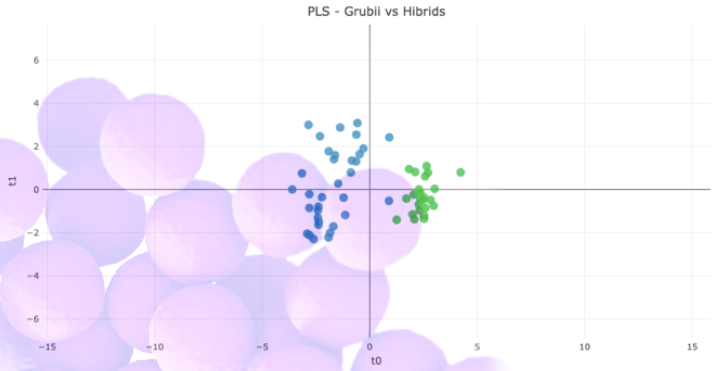

Figure 2. A PLS-2D clustering as a result of the second step peak matrix when it is analyzed in a PLS module that allow classify var. grubii and var. hybrid.

Results

Once the two peak matrices are done and validated, a PLS-DA can be applied to check the suitability of the developed methodology. Categories are set for each group studied. Additionally, we used a k-fold cross validation to verify the practical utility of the method.

For the first peak matrix the resulting the analysis allowed the correct discrimination of *C. neoformans* var. *neoformans* form the other two varieties in 100% of cases. Besides, using the second peak matrix with 10-fold cross validation, *C. neoformans* var. *grubii* could be reliably separated from hybrids in 96.55% of the cases.

In this point of the process we can be sure that the peak matrices done are reliable to identify correctly these samples. Then we convert this result into a *prediction base* to evaluate blind samples subsequently.

Conclusion

The results obtained show that this method is reliable for the identification of Cryptococcus varieties if the selected masses are representatives and optimum for the classification. This workflow can be applied to other samples of different species by concluding in the generation of a good prediction base that can allow the identifications of the new blind samples.

References

1. Implementation of MALDI-TOF Mass Spectrometry and Peak Analysis: Application to the Discrimination of Cryptococcus neoformans Species Complex and Their Interspecies Hybrids. *M.E. Zvezdanova, M.J. Arroyo, G. Méndez, J.V. Guinea Ortega, L. Mancera, P. Muñoz, B. Rodríguez-Sánchez, P. Escribano*. Journal of Fungi, 6, 3030, 2020.