

Merging FT-IR and NGS for Simultaneous Phenotypic and Genotypic Identification of Pathogenic Candida Species

Laura Corte¹, Claudia Colabella¹, Luca Roscini¹, Volha Shapaval², Achim Kohler², Valeria Tafintseva², Carlo Tascini³ and Gianluigi Cardinali^{1,4}

¹ Department of Pharmaceutical Sciences - Microbiology, University of Perugia, Perugia (Italy)

² Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway

³ Azienda Ospedaliera dei Colli - Ospedale Cotugno, Napoli, Italy

⁴ CEMIN, Centre of Excellence on Nanostructured Innovative Materials - Department of Chemistry, Biology and Biotechnology - University of Perugia, Perugia, Italy

The rapid and accurate identification of pathogen yeast species is crucial for clinical diagnosis due to the high level of mortality and morbidity induced, even after antifungal therapy. For this purpose, new rapid, high-throughput and reliable identification methods are required.

We described a combined approach based on two high-throughput techniques in order to improve the identification of pathogenic yeast strains. Next Generation Sequencing (NGS) of ITS and D1/D2 LSU marker regions together with Fourier Transform Infrared Spectroscopy (FTIR) were applied to identify 256 strains belonging to *Candida* genus isolated in nosocomial environments. Multivariate data analysis (MVA) was carried out on NGS and FT-IR data-sets, separately. Strains of *Candida albicans*, *C. parapsilosis*, *C. glabrata* and *C. tropicalis*, were identified with high-throughput NGS sequencing of ITS and LSU markers and then with FTIR.

Inter- and intra-species variability was investigated by Consensus Principal Component Analysis (CPCA) which combines high-dimensional data of the two complementary analytical approaches in concatenated PCA blocks normalized to the same weight. The total percentage of correct identification reached around 97.4% for *C. albicans* and 74% for *C. parapsilosis* while the other two species showed lower identification rates. Results suggested that the identification success increases with the increasing number of strains actually used in the PLS analysis.

The absence of reliable FT-IR libraries in the current scenario is the major limitation in FTIR-based identification of strains, although this metabolomics fingerprint represents a valid and affordable aid to rapid and high-throughput to clinical diagnosis. According to our data, FTIR libraries should include some tens of certified strains per species, possibly over 50, deriving from diverse sources and collected over an extensive time period. This implies a multidisciplinary effort of specialists working in strain isolation and maintenance, molecular taxonomy, FTIR technique and chemo-metrics, data management and data basing.