Similarity Measurement of Electrophoresis Strands for Fungi Fingerprinting

I.-O. Stathopoulou and G.A. Tsihrintzis

Department of Informatics

University of Piraeus, Piraeus 185 34, GREECE

Complete Address: Karaoli & Dimitriou 80, Piraeus, 18534, Greece

Phone: (+30) 210-4142322 Fax: (+30) 210-4142264 E mail: {iostath, geoatsi}@unipi.gr

G. Gaitanis, K. Kollia and A. Velegraki

Mycology Reference Laboratory, Department of Microbiology, Medical School

University of Athens, Athens, GREECE

Abstract – **Yeasts thrive in every environment on earth, understanding them, may help us to treat more rapidly and accurately (identification and elimination) the human diseases or design customized treatments depending the patient's needs. Clavispora lusitaniae is haploid yeast with a sexual cycle. It is an emerging opportunistic pathogen and an ongoing clinical problem because the efficacy of amphotericin B chemotherapy in cases of C. lusitaniae candidaemia is debatable. Also, Malassezia yeasts are members of the normal human skin flora, agents of skin disorders, which affect millions of patients worldwide, and systemic infections in subgroups of hospitalized severely immunocompromised patients. In previous work, we developed clustering and classification algorithms for processing images of strands of C. lusitaniae yeasts obtained via electrophoresis with the purpose of increasing the credibility of the analysis of molecular epidemiology data, facilitating and partially automating DNA fingerprinting processes. The algorithms consist of combinations of contrast and edge enhancement, segmentation, and adaptive filtering image processing techniques, which have been found to boost significantly the detection of electrophoretically separated chromosomes. Clustering and classification are effected using similarity measures (and corresponding dendrograms) based on prototype strands objectively defined as molecular weight size bands. In this paper, we improve our algorithm [1] and further testing it in the classification task of Malassezia's yeast electrophoresis images.**

Keywords - **Bioinformatics, Similarity measurement, clustering, classification, image processing, DNA profiling**

I. INTRODUCTION

DNA fingerprinting has established itself as an efficient and highly accurate means of determining identities and relationships. DNA profiling, as the process is more appropriately called, involves the visualization of special segments of the human genome, which are unique to each individual. This process can allow us to decode and further understand the fairly complex way the organisms live and interfere with each other.

Since the scientists were able to decode the DNA sequence, many sciences, which use DNA data, have been developed. Specifically, we have benefits in Medicine, because by decoding some important fungal species, medicines can treat (identify and eliminate) human fungal diseases (e.g. Candida and Malassezia species) more rapidly and accurately or design customized treatments, depending on the patient's DNA sequence [2-9]. Also, we can develop programs for educating future scientists and doctors in order to recognize some important fungal species more rapidly and accurately.

Moreover, we have benefits in Energy and Environmental Sector and in Agriculture and Food Production [10-17], as by understanding some microorganisms and microbes that thrive in every environment on earth, may allow us to exploit their abilities, in order to clean toxic wastes or make crops more resistant to water stress conditions and diseases and increase their reproductive capacity phases or, simply use them as starters for the production of specific foods.

Finally, the gain regarding the Bioanthropology section is very important, as we can understand human lineage or explore migration patterns through time and understand how they affect the different human species today. Also, when it comes to Human Identification [18], we can use DNA fingerprinting to identify potential criminals or kinships and victims of natural or human error disasters

The present work is the outcome of a joint effort of the Mycology Laboratory of the Department of Microbiology of Medical School of University of Athens and the Department

Manuscript received June 30, 2006.

of Informatics of University of Piraeus. Our aim is to develop a clustering and classification system for processing images of fungal strands, so as to increase the credibility, facilitate and partially automate fungal DNA fingerprinting process, which until now was done manually. Some of the challenges that have to be addressed in developing such a system arise from the fact that depending the method and the conditions used during electrophoresis, the quality of the image may vary significantly, so an algorithm capable to confront different image qualities, is needed

In this work, we used digital images of Clavispora lusitaniae and Malassezia yeasts as paradigm in order to apply our classification algorithm. Clavispora lusitaniae belongs to Candida genus and its infections make only 1% of nosocomial Candida infections, but have a poor prognosis for candidaemia, endocarditis, osteomyelitis and meningitis. Poor prognosis is attributed to a doubtful clinical response to amphotericin B, despite apparent in vitro susceptibility to the drug. Amphotericin B treatment failure is ascribed to mutations of the originally susceptible infecting strain, and to phenotypic/genotypic switching mechanisms selecting for resistant subpopulations during therapy. It is therefore imperative to timely recognize epidemics in the hospital (particularly in Intencive Care Units-ICU) so as to apply suitable management and control strategies, such as appropriate treatment regimens and/or timely disinfection of hospital Wards.

Malassezia yeasts are members of the normal human skin flora, agents of skin disorders and systemic infections in subgroups of hospitalized severely immunocompromised patients. As the incidence of yeasts in deep-seated infections continues to increase in proportion to the growing number of immunocompromised, cancer and postoperative patients, standardized procedures in assessing whether specific strains are responsible for hospital epidemics can become purposeful to clinical practice, patient welfare and health economics.

Malassezia yeasts apart from causing pityriasis versicolor (PV) are also implicated in the pathogenesis of various dermatoses with universal distribution as is atopic dermatitis and seborrheic dermatitis, while recent information suggests their involvement in psoriasis. All these diseases affect millions of patients worldwide. Published epidemiological data suggest geographical variations in the rate of the isolated species and molecular typing methods have been developed to evaluate distribution of different Malassezia subtypes within a given disease spectrum, yet without successfully relating molecular types with pathogenesis. Moreover, studies involving the contribution of different allergens in the pathogenesis of AD have taken place while different Malassezia species have been scrutinized for the existence and potential polymorphisms in sequences coding for the first major allergenic protein (Mala s 1) [19]. Therefore, development of a reliable system for recognizing clusters of Malassezia molecular subtypes has a twofold purpose: (a) to timely identify potentially fatal hospital epidemics and (b) to differentiate among geographically distinct pathogenic strains, hence supporting associations between specific molecular types and types of disease.

In Section 2 the clinical strain isolation and DNA extraction procedures of the fungal species (material) are presented, followed by the methods used during electrophoresis. In Sections 3 and 4, the Classification Algorithm and initial results are presented, respectively. Finally, in Section 5 conclusions drawn and, in Section 6, future work is outlined.

II. MATERIAL AND METHODS

A. *Strains*

C. lusitaniae is fungus belonging to the yeasts and causing infections that make only 1% of nosocomial Candida infections, but have a poor prognosis for candidaemia, endocarditis, osteomyelitis and meningitis. Poor prognosis is attributed to a doubtful clinical response to amphotericin B, despite apparent in vitro susceptibility to the drug. Amphotericin B treatment failure is ascribed to mutations of the originally susceptible infecting strain, and to phenotypic/genotypic switching mechanisms selecting for resistant subpopulations during therapy. It is therefore imperative to timely recognize mutations, by evaluating diverse fingerprints, epidemics in the hospital, particularly in Intencive Care Units-ICU, so as to apply suitable management and control strategies, such as appropriate treatment regimens and/or timely disinfection of hospital Wards.

Ten C. lusitaniae clinical strains, the Type strain CBS 6936 (mating type h+) and the reference strain CBS 5094 (mating type h-) were used in the study. The 10 C. lusitaniae strains were collected between 1998 and 2001 from an equal number of a seemingly, in epidemiological terms, unrelated patient cohort. Isolates originated from: bloodstream infections (6), pulmonary infection (1), oral lesions of patients undergoing chemo-radiotherapy for head and neck tumours (2), and from a patient with vaginal infection (1) following hysterectomy. All strains were identified by the API 32C system (BioMerieux, Marsy l' Etoil, France). Complementary biochemical and physiological tests were also performed for accurate characterization of the C. lusitaniae strains [20].

A total of 109 positive cultures for Malassezia from patients with PV (n=71); SD (n= 38) were obtained from North European, South and South-East European resident patients, identified to species level, analyzed by polymerase chain reaction (PCR) fingerprinting and included in the analysis.

B. DNA extraction PCR fingerprinting and Electrophoresis

Genomic DNA for PCR fingerprinting was extracted as follows: Clavispora lusitaniae strains were cultured on Sabouraud dextrose agar (Difco, Detroid, MI, USA) for 48 h at 30oC. One loopful of a standard inoculation loop (Greiner, GmbH, Germany, SAL 10-3) from each culture was transferred into 1.5 ml microcentrifuge tubes containing 500 μl lysis buffer (200 mM Tris-HCl, pH 8, 250 mM NaCl, 25 mM EDTA, 0.5% sodium dodecyl sulfate) (all from Sigma) and 6-8 glass beads 1.1-1.2 mm in diameter (Sherwood, St Louis, USA). The tubes were subsequently vortexed for 4 min and DNA was phenol:chloroform extracted as described before [21].

Pathological skin scales from patients with PV and AD were inoculated in modified Dixon's medium [3.6% yeast extract, 0.6% mycological peptone, 1% agar No1, 2% bile salts, 1% Tween 40, 0.2% glycerol (All from: OXOID, Basingstoke, United Kingdom)] supplemented with cycloheximide (0.02%) and chlorampenicol (0.005%) and were incubated in 9 cm diameter Petri dishes at 32oC for 2 weeks. DNA from each population of Malassezia yeasts grown on this medium was extracted as described above.

The minisatellite specific oligonucleotide [5'-GAGGGTGGCGGTTCT-3'] [22] was used as a single primer to amplify inter-repeat DNA sequences of C. lusitaniae and Malassezia yeasts. The PCR reactions contained (in a total volume of 25 μl) 2.5 μl of DNA template, 1U Taq polymerase (Promega, WI, USA), 1.5 mM MgCl2, dNTPs (Clontech, CA, USA) at 1.5 mM each, and 150 pmol primer (Interactiva GmbH, Ulm, Germany). PCR was performed in Grade 40 Stratagene Robocycler (CA, USA) for 45 cycles at: 95oC 1 min, 42oC 1 min 30 sec, 72oC 1 min 30 sec, with an added final extension at 72oC for 5 min. The profiles were separated in a 1.8% standard agarose gel electrophoresed for 1.5 h in 0.5 TBE at 60 V and stained with ethidium bromide. Each strain was tested at five independent occasions to ensure reproducibility of the analysis.

C. The Marker

The molecular size marker used as a reference for a measure of bands sizes was the 100 base pair (bp) DNA ladder ranging from 100 - 1, 517 bp, corresponding to 25 – 45 ng of DNA mass. The marker is shown in Figure 1, where the bp value of each strand is depicted. This value depicts the migration speeds, during electrophoresis, of each band. The location of all the bands in a yeast strand differs, depending the family in which the yeast belongs. The bp value of the each strand is computed based on its relative position to the equivalent marker's band, which we use as a reference (reference strand).

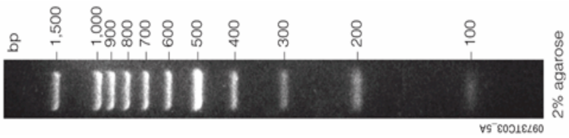

Fig. 1. The 100 base pair DNA marker

III. THE CLASSIFICATION SYSTEM

During electrophoresis session, a group of strands are being analyzed and a digital image is been produced, depicting this group of strands side by side. Usually this group consists of 14 or 15 strands; among them, there is the reference band, which is repeated 3 of 4 times. As we use this band to compute the total bp value of each strand, this repetition is needed for reasons of precision and accuracy, since, until now, the classification task was done manually and the detail is important.

Our classification system aims to automate this procedure by clustering each stand and computing the bp values based on the location of each band, relatively to the reference strand.

A. The Algorithm

In detail, the proposed classification algorithm works as follows:

1) We load the candidate image
2) The user selects the reference strands and the strands that she/he wants to be classified
3) We preprocess the image in order to enhance its quality and to render visible more bands. The image preprocessing step is very important for the successful classification of the strands
4) Each strand is scanned from top to bottom and the position and size of the parts of the image where the pixels are white (equal to 1) is computed
5) We compute the similarities based on these distances and the reference band
6) We produce a dendrogram and present the classification result

B. Selection of the corresponding strands

After loading the digital image of the group of strands, the user must select the reference strands and the strands that she/he wants to be classified. This is done, because the reference strand is repeated randomly in the group of strands. The selection is shown in Figure 2, where it is presented the digital image obtained by the electrophoresis of 15 strands, among them, the strands No 1, 6, 11 and 15 are the reference strands and the rest need to be classified

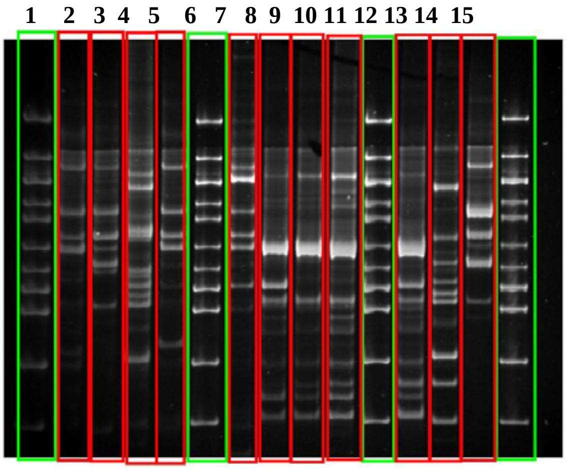

Fig. 2. The selected strands

C. Image Preprocessing

Depending on the method, the gel and the conditions used during electrophoresis, the quality of the image may vary significantly and a need arises for algorithms invariant to such image quality. Our approach includes the following ‘image preprocessing step’, which improves the quality of digital strand images and enhances the visibility and discrimination of bands:

7) Load the acquired image and covert it to grayscale (if necessary)
8) Compute a 3-by-3 unsharp/contrast enhancement filter from the negative of the Laplacian filter with parameter 0.2 and apply it to the input image
9) Adjust image intensity values, so that 1% of data is saturated at low and high intensities of the input image. This further enhances the contrast in the resulting image.
10) Perform two-dimensional median filtering to reduce noise and preserve edges.
11) We apply k-means clustering in the preprocessed image, formatting 2 clusters and color each cluster with the average color
12) We convert the clustered image to binary

The aforementioned algorithm improves the image quality significantly, but there exists a limit to the preprocessing improvement, since further preprocessing may deform and tamper with those bands that are already intense.

The k-means clustering step, after the preprocessing, is very important, as it discriminates bands that they weren’t so intense. So in the resulting image, we can observe bands that they weren’t so visible to the naked eye. Typical results after image preprocessing are given in the following Figure 3, where the resulting image after preprocessing, can be compared to the resulting binary image without preprocessing.

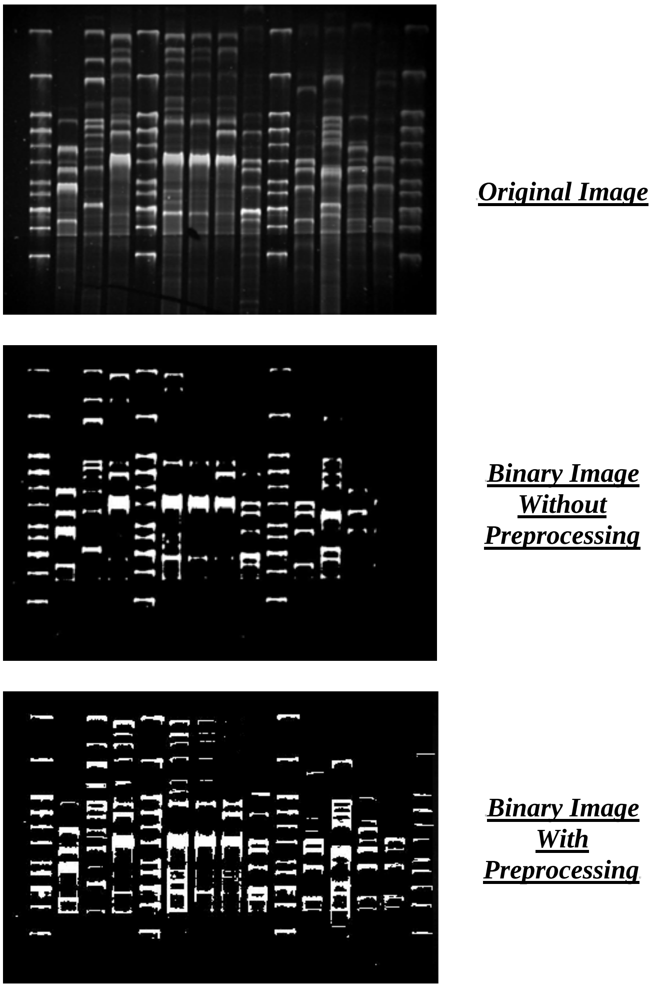
Original Image
Binary Image Without Preprocessing
Binary Image With Preprocessing

Fig. 3. Image Preprocessing Results

IV. RESULTS

In the following Figure 4, we see typical similarity dendrograms produced by our system. In these dendrograms, strand similarity increases (decreases) with dendrogram depth (height) and markers are positioned as the rightmost strand.

Based on the observations made by the Mycology Laboratory, our classification system succeeded in classifying the strands according to the family which they belong. This procedure can greatly help the DNA fingerprinting process.

Our image analysis procedure clustered epidemiologically linked C. lusitaniae and Malassezia yeast isolates in the same group. This timely identified C. lusitaniae life-threatening infection epidemics in immunocompromised patients in the hospital. Regarding from different European geographical regions the analysis clustered pathogenic isolates in distinct subgroups indicating a georgraphical gradient among Malassezia yeasts, which is recorded for the first time.

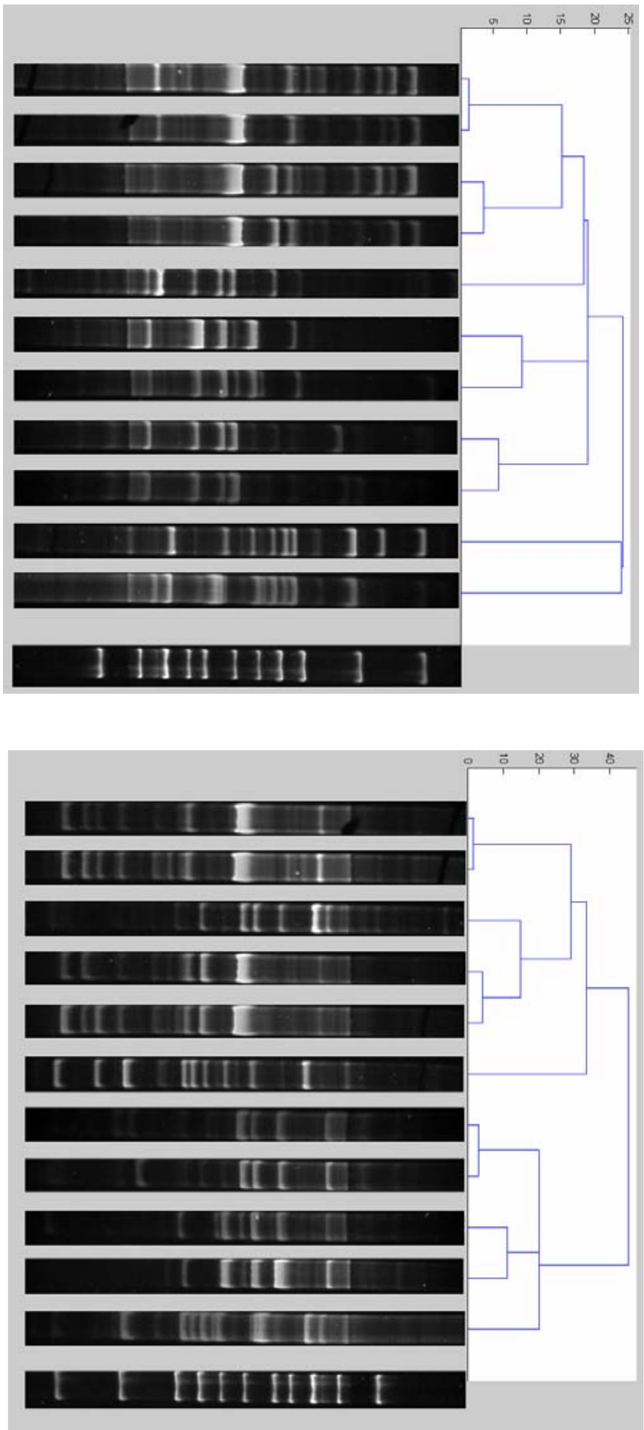

Fig. 4. Strand Similarity Dendrograms

V. CONCLUSIONS

Automated image data processing would greatly improve handling of a large volume of fingerprint data by the Mycology Laboratory. Also, the rapid analysis and reliable assessment of data facilitates timely implementation of management measures to control hospital epidemics. This, apart from saving human lives, also contributes to substantial savings in hospital and treatment costs. The developed system is of low cost and user-friendly. Its simplicity meets the criteria for use in familiarizing Science and Medical undergraduate and graduate students with bioinformatics during practical sessions. As mentioned in Section 3.2., image preprocessing is a prerequisite in order to improve the image quality. It is possible that in cases where more detail is needed, such as strands with index of similarity, further preprocessing may be required in parts of the image.

VI. FUTURE WORK

This work will be expanded in the following three directions: (a) we will develop the graphical user interface of our system and making it straightforward to the user. (b) After completion of the previous step, users (Biological sciences and Medical students, academic staff and Health professionals in collaboration with the Mycology Reference Laboratory) will test and evaluate the software in detail. (c) Taking into consideration the outcome of the evaluation, we will extend the software further, so as to cover cases of images with serious registration problems, such as very low quality, long exposure time, alignment errors, etc.

REFERENCES

[1] Stathopoulou, I.-O. and Tsihrintzis, G.A., “Clustering and Classification of Electrophoresis strands for fungi fingerprinting”, Proceedings of the 7th International Workshop on Mathematical Methods in Scattering Theory and Biomedical Engineering , September 17-21, 2005

[2] Qing-Yin Zeng, Asa Rasmunson-Lestander and Xiao-Ru Wang, Extensive set of mitochondrial LSU rDNA-based oligonucleotide probes for the detection of common airborne fungi, FEMS Microbiology Letters, 79-87, 237 (2004)

[3] A. Espinel-Ingroff, J.A. Vazquez, D. Boikov and M.A. Pfaller, Evaluation of DNA-based Typing Procedures for Strain Categorization of Candida spp., Diagnosis of Microbiological Infections, 231-239, 33 (1999)

[4] Mamie Hui, Margaret Ip, Paul K.S. Chan, Miu Ling Chin, Augustine F.B. Chen, Rapid identification of medically important Candida to species level by polymerase chain reaction and single-strand conformational polymorphism, Diagnostic Microbiology and Infectious Disease,95-99, 38(2000)

[5] Stewart Scherer and David A. Stevens, A Candida albicans dispersed, repeated gene family and its epidemiologic applications, Proceedings Natl. Acad. Sci. USA, 1452-1456, 88(1988)

[6] B.A. Lasker, G.F. Carle, G.S. Kobayashi and G. Medoff, Comparison of the separation of Candida albicans chromosome-sized DNA by pulsed-field gel electrophoresis techniques, Nucleic Acids Res., 3783–3793, 17(1989)

[7] Aarturo Lopez, Noel Xamena, Ricard Marcos and Antonia Velasquez, Germline genomicinstability in PCNA mutants of Drosophila: DNA fingerprinting and microsatellite analysis, Mutation Research, 253-265, 570 (2005)

[8] W.G. Merz, Carla Connelly and Philip Hieter, Variation of Electrophoretic Karyotypes among Clinical Isolates of Candida albicans, Journal of Clinical Microbiology, 842-845, 1988

[9] Aristea Velegraki, Manousos E. Kambouris, George Skiniotis, Marianna Savala, Angeliki Mitroussia-Ziouva and Nicholas J. Legakis, Identification of medically significant fungal genera by

polymerase chain reaction followed by restriction enzyme analysis, FEMS Immunology and Medical Microbiology, 3003-312, 23 (1999)

[10] E. Yergeau, M. Filion, V. Vujanovic and M. St-Arnaud, A PCR-denaturing gradient gel electrophoresis approach to assess Fusarium diversity in asparagus, Journal of Microbiological Methods, 143-154, 60(2005)

[11] Renske Landeweert, Christiaan Veenman, Thom W. Kuyper, Hannu Fritze, Karel Wernars and Eric Smit, Quantification of ectomycorrhizal mycelium in soil by real-time PCR compared to conventional quantification techniques, FEMS Microbiology Ecology, 183-192, 45 (2003)

[12] Timothy R. Dean, Barbara Roop, Doris Betancount and Mark Y. Menetrez, A simple multiplex polymerase chain reaction assay for the identification of four environmentally relevant fungal contaminants, Journal of Microbiological Methods, 9-16, 61(2005)

[13] S. Arancia, S. Sandini, A. Cassone, F. De Bernardis and R. La Valle, Construction and use of PCR primers from a 65 kDa mannoprotein gene for identification of C. albicans, Molecular and Cellular Probes, 171-175, 18 (2004)

[14] Lia Rossetti and Giorgio Giraffa, Rapid Identification of dairy lactic acid bacteria by M13-generated, RAPD-PCR fingerprint databases, Journal of Microbiological Methods, in-press, (2005)

[15] W.K. Ma, S.D. S iciliano and J.J. Germida, A PCR-DGGE method for detecting arbuscular mycorrhizal fungi in cultivated soils, Soil Biology & Biochemistry, in-press, 2005

[16] F. Cappa and P.S. Cocconcelli, Identification of fungi from dairy products by means of 18S rRNA analysis, International Journal of Food Microbiology, 157-160, 69 (2001)

[17] Fabio Faria da Mota, Eliane Aparecida Gomes, Edilson Paiva and Lucy Seldin, Assesment of the diversity of Paenibacillus species in environmental samples by a novel rpo B-based PCR-DGGE method, FEMS Microbiology Ecology, in-press, (2005)

[18] Kimberly G. Nugent and Barry J. Saville, Forensic analysis of hallucinogenic fungi: a DNA-based approach, Forensic Science International, 147-157, 140 (2004)

[19] Gaitanis G, Menounos P, Katsambas A and A. Velegraki. Detection and mutation screening of Malassezia sympodialis sequences coding for the Mal s 1 allergen implicated in atopic dermatitis. J. Inv. Dermatol..2003;121(6):1559-60.

[20] Lachance MA, Phaff HJ. Descriptions of teleomorphic ascomycetous genera and species. Clavispora-Rodriguez de Miranda. In: Kurtzman CP, Fell JW. Eds. The Yeasts, a taxonomic study, 4th edn. Amsterdam: Elsevier Science Publishers, 1998:148-52.

[21] Velegraki A, Kambouris M., Skiniotis G, et al. Identification of medically significant fungal genera with PCR followed by restriction enzyme analysis. FEMS Immunol Med Microbiol 1999; 23: 303-12

[22] Meyer W, Mitchell TG. Polymerase chain reaction fingerprinting in fungi using sinle primers specific to minisatellites and simple repetitive DNA sequences: strain variation in Cryptococcus neoformans. Electrophoresis 1995; 16: 1648-165