Sequence basis of the serotyping and subtyping efficiency of URA-A/B PCR in *Cryptococcus neoformans* and indications for one-way multilocus allelic shift

Toukas Dimitrios *[1], Kambouris Manousos E.*[2], Gaitanis George[1,3], Menounos Panagiotis G.[4] and Velegraki Aristea[1]

1. Dept of Medical Laboratories, Fcty of Health and Caring Professions, Technological Educational Institute of Athens, Ag Spyridonos st., 12210 Athens, GREECE.
2. Mycology Laboratory, Medical School, National & Kapodistrian University of Athens 75-77 Mikras Asias st., 11527 Athens, GREECE
3. Dermatology and Venereology Department, Medical School, University of Ioannina, S. Niarchou Av 45110, Ioannina Greece
4. Molecular Biology Laboratory, Military School of Nursing, Athens, GREECE

* Authors of equal contribution

Abstract

The genetic basis of the molecular subtyping of *Cryptococcus neoformans* isolates, performed by PCR-RFLP and PCR-SSCP methodologies led through sequencing and sequence alignment to the detection of Single Nucleotide Polymorphisms (SNP) with consistent allelic status and a clear direction of allelic shifting within and among the accepted subspecies taxa (varieties and serotypes). The consensus sequence allelic status of the subtyped and serotyped isolates points to the possibility of serotype transition accomplished through a gradual allelic shift. The direction of the shift is from Serotype A to Serotype AD to Serotype D, although it is unknown whether these quantal steps can be taken by the same isolate.

Key words

Cryptococcus neoformans, URA5 gene, PCR, SSCP, Serotype AD, SNP, Loss-Of-Heterozygosity (LOH)

Introduction

Cryptococcosis is a serious infection of immunocompromised patients (8, 27) and is considered an AIDS-defining infection (17, 28). Fungal meningitis, pulmonary cryptococcosis and brain cryptococcoma have high morbidity (19). Furthermore, members of the genus have a widespread environmental distribution (5, 9, 12, 13), infect the host through inhalation (15). They are carried in long distances by pigeons (18) which abound in both urban and rural environments, and occur in abundance on the leaves and bark of eucalyptus trees (5), which are planted in rural and urban areas to act as natural insect repellants. Eukalyptus leaves are also used in alleviating respiratory tract discomfort. The species Cryptococcus neoformans is able to propagate at body temperatures, and has by far the highest clinical significance (3, 20). Its strains are diverse in biochemical and clinical properties, which led initially to varieties and serotypes being introduced in its taxonomy (2, 4) and subsequently to major taxonomic revisions (14). The diverse clinical properties and also the broad geographical distribution of the yeast led in the development of serotype-defining methodologies applicable to minimum specimen quantities for use with clinical samples, providing rapid results. This initiated PCR-based methodologies for the detection of Cryptococcus (23), identification to species level (24) and serotyping (25). Since the target URA5 gene sequences are neither variety- nor serotype-defining, yet, showed some heterogeneity, as SSCP discriminated two subtypes for each homogenous serotype (26), we investigated the molecular genetic basis of this subspecific discrimination in order to define the applicability using a single PCR amplicon for both identification and typing. Such a possibility would allow rapid and low cost typing since the amplicon is already obtained by the identification procedure, thus saving time, labour, consumables and eliminating a second DNA extraction step. Moreover, establishing a SNP-dependent genetic basis, would allow mapping and characterization of these SNPs, which might permit selection of informative restriction sites for further or alternative RFLP analysis.

Materials & Methods

Strains

TABLE 1. **Strains used as serotyped by molecular methodology and subtyped (where possible) by SSCP.**

Strain No	12	42	17	23	26	35	49	19	45	61	69	77	81	82
Subtype	AI	AII	DI	DII	AIDII	AID	AIDI	DII	DI	DII	DII	DII	DII	DII

DNA extraction

Single colonies of *Cryptococcus neoformans* (*C.n*) var. *neoformans* clinical isolates of serotypes A, D and AD, were used on multiple independent experimental occasions. Serotypes were determined by using the PCR-RFLP methodology described previously (**25**). Genomic DNA for genotyping and for PCR was extracted as described before (**24, 25**). Briefly, all yeast strains were cultured on Sabouraud dextrose agar (Difco, Detroid, MI, USA) for 48 h at 30°C. 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 St. Louis, MS, USA) 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 extracted once with phenol: chloroform : isoamyl alcohol (25 : 24 : 1 v : v : v) and once with chloroform : isoamyl alcohol (24 : 1 v : v). Nucleic acids were precipitated by adding an equal volume of cold isopropanol (Merck, Darmstadt, Germany). After centrifugation for 10 min at 8,000 g the pellet was washed in 500 µl wash buffer (76% ethanol, 10 mM ammonium acetate) and allowed to stand at room temperature for 15 min. The suspension was centrifuged at 8,000 g for 5 min at room temperature and the pellet was allowed to dry for 2-4 min. DNA was then washed in 70% ethanol and collected by centrifugation as before. The pellet was dried and resuspended, depending upon yield, in 35-50 µl sterile distilled water. (**24**).

PCR primers and reaction conditions

PCR reactions were performed as described previously (**25**) in 100 µl volume containing 10 µl suspension of template DNA, 1.5 mM $MgCl_2$, 0.4 mM dNTPs each (CLONTECH, Palo Alto, CA, USA), 2U *Taq* (Promega, Madison, WI, USA), 10 µl of 10× reaction buffer (10mM Tris-HCl, pH 9.0, 50mM KCl, 0.1% Triton®X-100) and 50 pmoles of each primer. The forward and reverse primers *URA*5-A/B (Interactiva Biotechnologies, Ulm, Germany) were derived from *C. neoformans URA*5 gene, which generate an amplification product of 345 bp (position 1145 – 1490, GenBank accession no. M34606) and have been reported to amplify only *C. neoformans* sequences **(22)**.

*URA*5 A: 5'-ACG GTG AGG GCG GTA CTA TG-3'
*URA*5 B: 5'-AAG ACC TCT GAA CAC CGT AC-3'
Amplification was performed in an authorized RoboCycler Grandient 96 Hot Top Combo (Stratagene, LaJolla, LA, USA) (Ericomp, San Diego, CA, USA) at 40 cycles, 1min at 94°C, 1min 30 sec at 63° C and 1min at 72°C. The products were visualised in 2.5% agarose (Sigma St. Louis, MS, USA) prepared in 0.5X TBE, stained with ethidium bromide and run for 30 min at 100V for assessment of the amplicon yield.

PCR - SSCP analysis and sequencing

Non-radioactive SSCP analysis of the 345 bp amplification product was performed as follows: After purification with the DyeEx 2.0 spin kit-50 (Quiagen GmbH, Hilden, Germany), 5 µl of each amplification product were added to 5 µl of 95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol (all from Sigma, St. Louis, MS, USA)], heated to 95° C for 15 min and quenched on ice for 20 min to achieve almost complete denaturation. Each sample was then loaded on a 10% non-denaturing polyacrylamide gel, 29 w/w acrylamide :1,25 w/w bisacrylamide (Sigma, St. Louis, MS, USA) :5% Glycerol (Merck Darmstadt, Germany). Electrophoresis was performed at 380 V, 3 A for 22 h using the Marcophor (Pharmacia, Uppsala, Sweden) running apparatus at 12°C (P-Selecta Frigiterm 6000382, J.P Selecta S.A., Abrera, Spain) and visualised by 0,5% silver nitrate (BioRad, Munich, Germany) staining.
To cross-examine and confirm the SSCP results, all amplification products were sequenced by the automated ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and the sequences were aligned using the "Bioedit Sequence Alignment Editor" program.
Consensus sequences were determined by comparing the chromograms of the forward and reverse

primers, wherever available, for the homogenous serotypes (namely A and D). Then, the consensus sequences were aligned and compared in four distinct levels of importance. First were compared sequences of the same serotype, subsequently the sequences of both homogenous serotypes, then the sequences of all the homogenous serotypes (A and D) were compared against the heterologous serotype (AD) strains and last came the comparison to GenBank (www.ncbi.nih) retrieved sequences where the corresponding serotype **was indicated by the depositors**.

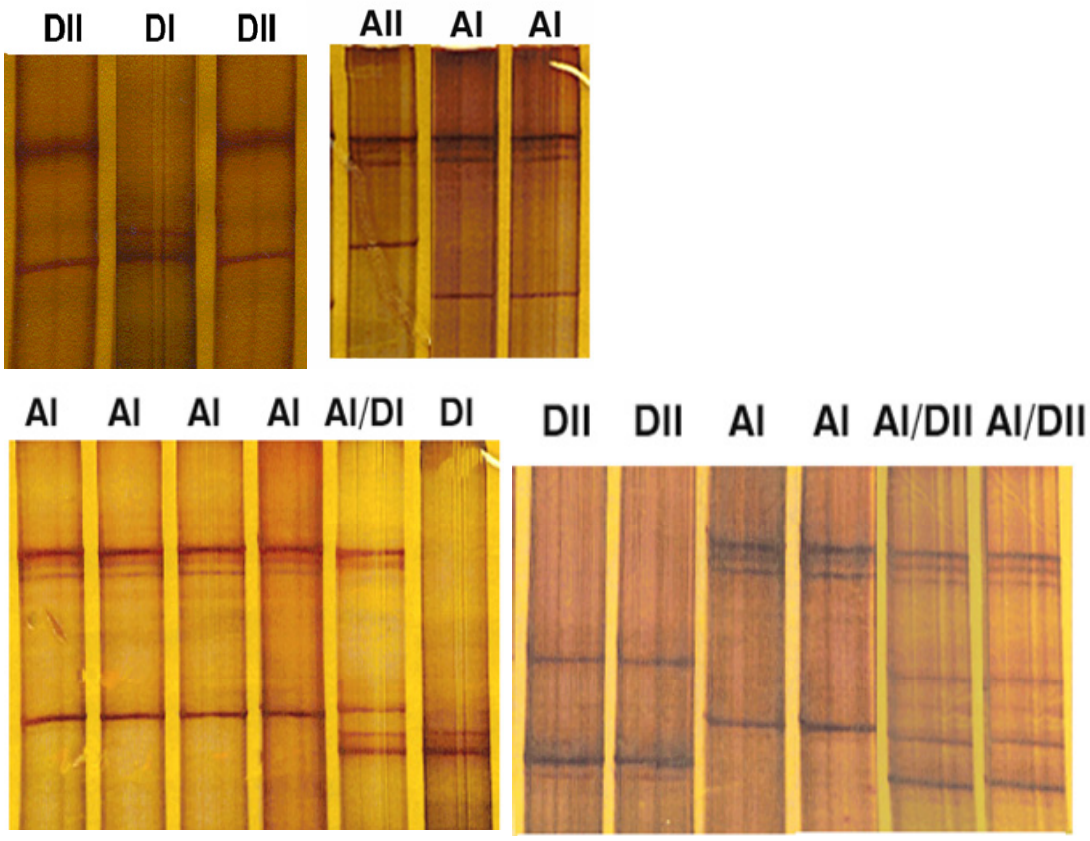

FIGURE 1: **Comparative SSCP patterns of *C. neoformans* Serotypes A, D, AD subtypes.**

Results

The A and D serotypes differed by one nonsense transition substitution in Residue No 30 and position 3 of the 10th codon. A transition heterozygocity symbol (Y or R, depending on the primer's vector) in the base calling process constitutes thus a safe indicator for AD serotype.

The A_1 and A_2 subtypes where found to differ by one nonsense transition substitution in Residue No 75 and position 3 of the 25th codon. This biallelic SNP's status differentiated A_2 from A_1, D_1 and D_2, all of which were clustered together as they shared the same allele. A transition heterozygocity symbol (Y or R, depending on the primer's vector) in the base calling process constitutes thus a safe indicator for A_2D_x subtype, whereas a homozygotic A or T (depending on the primer's vector) indicates a A_1D_x subtype.
The comparison between D_1 and D_2 subtypes revealed a strikingly different picture. The most important finding is the presence of 13 nonsense transition mutations (biallelic SNPs), all being in-frame at the 3rd residue of the respective codons (residues 60, 69, 87, 117, 132, 135, 228, 249, 252, 255, 258, 270, 279). The 13 transitions observed between D_1 and D_2 cluster D_1 along with A_1 and A_2 and apart from D_2. A nonsense transversion mutation at Residue 319 (1st position of 107th codon) also clusters D_2 apart from A_1, A_2 and D_1 which shared the same allele. The tranversion heterozygocity symbol (K or M, depending on the primer's vector) in the base calling process constitutes thus a safe indicator for A_xD_2 subtype, whereas a homozygotic C or G (depending on the primer's vector) indicates a A_xD_1 subtype.

```
                  ..........10.....20.....3O.....40.....50.....60.....70.....75.....80
AI          ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCACCTCTCAAGGGACGAATCGTCATCATCGATGATGTTCTCACCTCTGGCAAG
AII         ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCACCTCTCAAGGGACGAATCGTCATCATCGATGATGTTCTCACCTCCGGCAAG
DI          ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCGCCTCTCAAGGGACGAATCGTCATCATCGATGATGTTCTCACCTCTGGCAAG
DII         ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCGCCTCTCAAGGGACGAATCGTCATCATCGACGATGTTCTCACCTCTGGCAAG
AIDII       ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCRCCTCTCAAGGGACGAATCGTCATCATCGACGATGTTCTCACCTCTGGCAAG
(35)        ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCRCCTCTCAAGGGACGAATCGTCATCATCGACGATGTTCTCACCTCTGGCAAG
(26)        ⟶     ACGTGAGGGCTTTACTATGGTCTGGTGCRCCTCTCAAGGGACGAATCGTCATCATCGAYGATGTTCTCACCTCTGGCAAG
(49)        ⟶     AGGTGAGGGCTTTACTATGGTCTGGTGCRCCTCTCAAGGGACGAATCGTCATCATCGAYGATGTTCTCACCTCTGGCAAG

                  ..........90.....100.....110.....120.....130.....140.....150.....160.....
AI          ⟶     GCCATTCGTGAAGCTATTGACATTCTCAAGGCCTCTCCTGAAGCGAAGCTCGTTGGAATTGTCCAGCTTGTCGACAGACAAG
AII         ⟶     GCCATTCGTGAAGCTATTGACATTCTCAAGGCCTCTCCTGAAGCGAAGCTCGTTGGAATTGTCCAGCTTGTCGACAGACAAG
DI          ⟶     GCCATTCGTGAAGCTATTGACATTCTCAAGGCCTCTCCTGAAGCGAAGCTCGTTGGAATTGTCCAGCTTGTCGACAGACAAG
DII         ⟶     GCCATCCGTGAAGCTATTGACATTCTCAAGGCCTCCCCTGAAGCGAAGCTTGTCGGAATTGTCCAGCTTGTCGACAGACAAG
AIDII       ⟶     GCCATCCGTGAAGCTATTGACATTCTCAAGGCCTCCCCTGAAGCGAAGCTTGTCGGAATTGTCCAGCTTGTCGACAGACAAG
(35)        ⟶     GCCATCCGTGAAGCTATTGACATTCTCAAGGCCTCYCCTGAAGCGAAGCTYGTYGGAATTGTCCAGCTTGTCGACAGACAAG
(26)        ⟶     GCCATYCGTGAAGCTATTGACATTCTCAAGGCCTCYCCTGAAGCGAAGCTYGTCGGAATTGTCCAGCTTGTCGACAGACAAG
(49)        ⟶     GCCATYCGTGAAGCTATTGACATTCTCAAGGCCTCYCCTGAAGCGAAGCTYGTYGGAATTGTCCAGCTTGTCGACAGACAAG

                  ..........170.....180.....190.....200.....210.....220.....228
AI          ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCT
AII         ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCT
DI          ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCC
DII         ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCT
AIDII       ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCT
(35)        ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCT
(26)        ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCY
(49)        ⟶     AGAAAGGCCAGAGCGGTAGCGGCAAGAGTACCGTACAGGAGGTTGAGGAAGAGTTCGGTGTGCCY
```

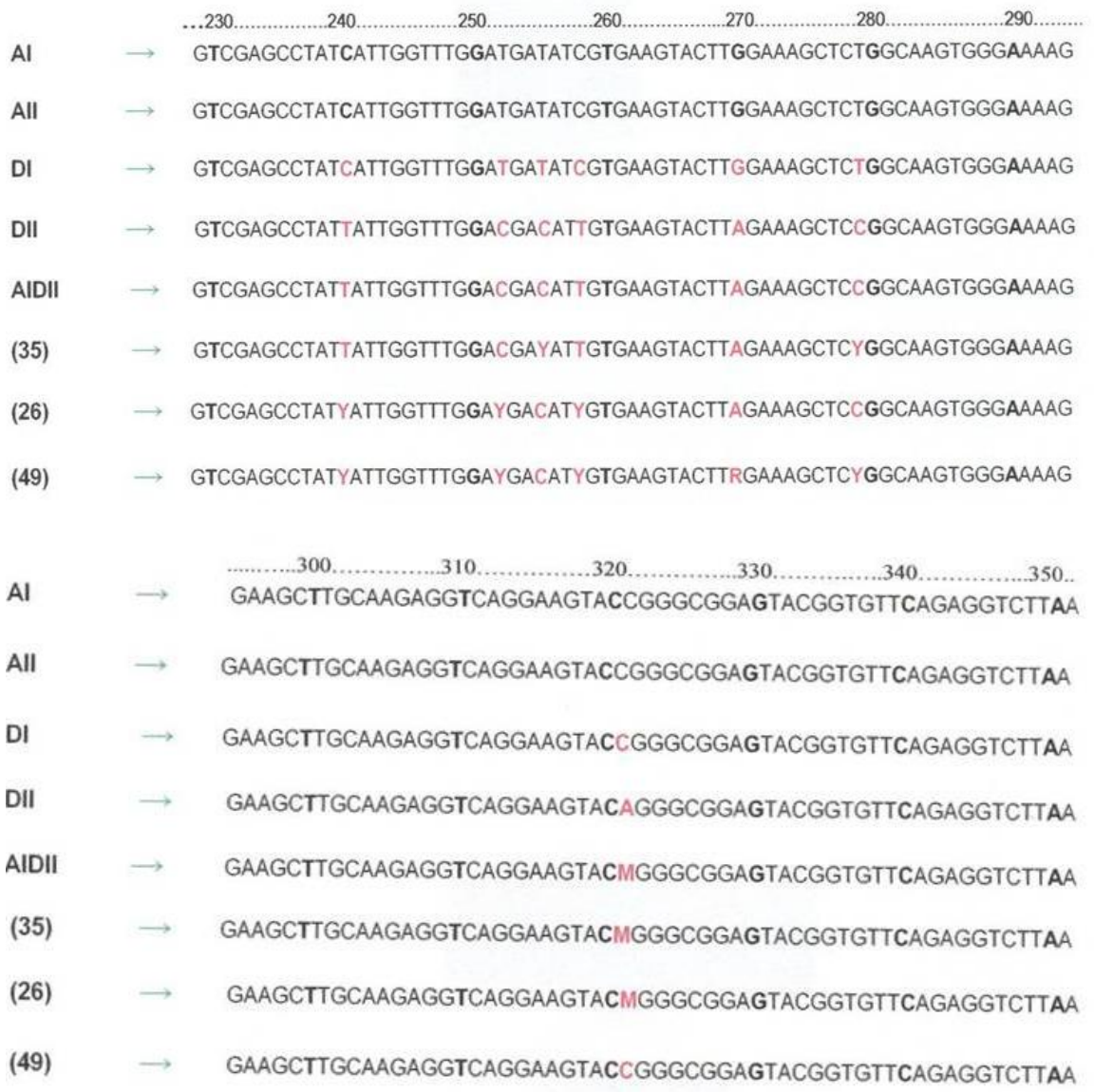

FIGURE 2: **Aligned sequence results of various subtypes (A_1, A_2, D_1, D_2, A_ID_{II}) and in-house AD strains (Strains 26, 35, 49). R=A/G, Y=C/T, M=A/C dimorphisms. Residues differentiating between D_1 and D_2 are in red, the ones differentiating between A_1 and A_2 are in green and the ones differentiating between Serotype A (A_1, A_2) and Serotype D (subtypes D_1. D_2) are in blue.**

Some 6 more D_2 strains plus our original D_2, strain 23 (included for control purposes) and one more D_1 strain plus our original D_1, strain 17 (included for control purposes) were subsequently sequenced to assess the extent and homogeneity of the haplotypes defining D_1 status. The 12 of the 13 observed transitions do cluster D_1 along with A_1 and A_2 and apart from D_2. The 13th, in Residue 69 is characteristic for only one D_2 strain, the No 23, while the other D_2 strains had the same allelic status at this position as D_1 and A_1, A_2 and is thus considered recent point mutation. To our amazement, both D_1 strains had fully shifted in all 13 informative positions from D_1 to D_2 i.e. the whole D-informative haplotype of 12 transitions and one transversion shifted from $A_1/A_2/D_1$ to D_2. The completeness of the shifting is absolute and there is no sequence variation, neither indel nor substitution, to distinguish the 7 D_2 strains from the 2 D_1 ones. This practically means that the Strtains 17 and 45 shifted to D_2 haplotype at some time after predated SSCP experiments were performed, as this had happened quite some time before the second batch of sequencing experiments.

TABLE 2: **Allelic status of informative residues in homogenous serotypes and selected AD-serotyped strains**

Residue No	St 35 A1Dx	Str 26 A1D2	St 49 A1D1	A1 Haplotype	D2 Haplotype
30	-	R	R	A	G
60	C	Y	Y	T	C
69	C	C	C	C	T
75	T	T	T	T	T
87	C	Y	Y	T	C
117	Y	Y	Y	T	C
132	Y	Y	Y	C	T
135	Y	C	Y	T	C
228	T	Y	Y	C	T
240	T	Y	Y	C	T
252	C	Y	Y	T	C
255	Y	C	C	T	C
258	T	Y	Y	C	T
270	A	A	R	G	A
279	Y	C	Y	T	C
319	-	M	C	C	A

Going to the third level of comparison, 3 of our own AD strains (Strains 26, 35, 49) were sequenced and compared to the individual A and D sequences and with SSCP results. The interesting finding is that they were supposedly A_1D_1 heterozygotes with observed diploidy by preceding SSCP, but the sequence alignment brought occasionally forth a A_1-D_2 pattern. The symbols $A_{1/2}$, $D_{1/2}$ were arbitrary introduced by us to describe SSCP subtypes within the serotypes (**26**). We kept the SSCP nomenclature for the homogenous serotypes and recharacterised the AD strains by sequence relevance. Sequencing has shown one of our AD strains (Strain 26) being A_1D_2. It bears R heterozygocity at Residue 30, the A_1 allele at Residue 75 and M heterozygocity at Residue 319. The second of our AD strains (Strain 49) was shown as A_1D_1. It bears R heterozygocity at Residue 30, the A_1 allele at Residue 75 and $A_1/A_2/D_1$ allele at Residue 319. The third AD strain (strain 35) cannot be properly subtyped by sequencing, because both Resdue 30 and Residue 319 are outside the sequenced area. As at Residue 75 it bears the A_1 allele it can be characterized A_1D_x despite the lack of the informative heterozygocity at Residue 30, due to the numerous heterozygocities scored at the 12 transition spots, which clearly indicate diploidy with presence of some D_2 alleles. SSCP had given clear-cut A_1D_1 patterns for all 3 strains at earlier time.

More intriguing was the fact that the expected heterozygosity was not universally present to all 12 transition spots/loci in any of the 3 strains: there were enough cases where homozygocity was observed in expected heterozygocity points, and these homozygous points were neither standard nor the same in the 3 strains. A weighted clustering of the 3 strains is possible, but no identical pairs of strains could be detected in the distribution of the heterozygosity and homozygocity observed on the 13 expected spots (12 transitions, 1 transversion). Even more intriguing is the fact that wherever a homozygocity was observed in any of the 13 positions in any of the 3 strains (a total of 16 homozygocities in 39 loci), it clearly implied a D_2 haplotype.

The transversion event at Residue 319, on the other hand, proved of more practical value. The chemical process of this type of mutation is more complicated and thus more stable. At strain 35, the Residue 319 lays outside the sequenced area, but at strain 26 there was a clear heterozygocity (implying A_XD_2) whereas at Strain 49 a clear homozygocity. The important fact is that it is the only SNP homozygocity event in the polymorphic positions not presenting the D2 allele. It thus clearly implies the A_1D_1 subtype. A SNP presenting such stability is considered the most informing, and possibly the only one necessary, for attesting D_2 status to a D-serotyped strain. Thus we characterized strain 49 as A_1D_1, in accordance to previous SSCP, and strain 26 A_1D_2 based on the allelic status of Residue 319, in contrast to previous SSCP. In the AD strains, the haplotype transforms progressively towards D_2. It is observed that only the A/D defining transition at Residue 30 and the D_1/D_2 defining transversion of Residue 319 do not shift to homogenous D_2 alleles and present stability.

Of extreme importance is the fact that Strain 49 retains its A_1D_1 characteristic allelic condition at Residue 319 despite the fact that all the other D_1/D_2 SNP positions have either transited to A_1D_2 heterozygocity, or (Residue 255) to homozygocity for D_2 allele. Contrarily, Strain 26, characterized A_1D_2 by the transversion, has an allelic status of the 12 SNPs highly/more weighted to D_2: there are 8 heterozygocities and 4 D_2 homozygocities. If a model of allele shifting is accepted, it somehow is asynchronous in the different positions and not needing a D_2 strain to initiate the shift through a LOH mechanism. A series of steps turn the SNP positions to A_1D_2 heterozygous and subsequently to D_2 homozygous. At some point, when the shift (teleologically entailing 12X2= 24 steps) reaches a threshold, the character of the D_1 chain changes to D_2, a change signaled by the transversion of Residue 319 to heterozygous status. Neither the threshold nor its nature is known, i.e. whether it entails a strictly predetermined sequence of steps or a less stringently predetermined one, with some degrees of freedom. Though, obviously some positions may conclude the transformation to full D_2 status with 2 transitional steps (e.g. Position 255) before the position 319 has transversed to A_1D_2 allelic condition. Thus, it is highly recommended that to determine the subtype of a *C. neoformans* var. *neoformans* strain, one examines equivalently the allelic status at 3 positions : Residue 30, for A/D/AD serotyping, Residue 75 for A_1/A_2 discrimination and Residue 319 for D_1/D_2 discrimination. Regrettably the lack of A_2D_x strains (if existent in nature) has not allowed examination of concurrence between allelic status at positions 75 ant 319 and possible dependence relations between the two and towards Position 30.

Unfortunately, the SSCP and sequencing experiments were not concomitant and thus SSCP results cannot authenticate the time variable, as between subcultures transitional events evidently take place, and perhaps more than one at a time in some cases.

The comparison of our own results to sequences deposited in the NCBI public database (GenBank) from serotyped strains showed some interesting points: First, all 6 A-serotyped strains [GenBank (www.ncbi.nih) Accession Nos AJ555621.1, AF248097.1/098.1/099.1/100.1/101.1] were of the A_1 subtype, by force of the single SNP at Position 75. Second, practically there were not AD serotypes, as at Residue 30 they were scored homozygously. Third, as far as AD serotypes are concerned, there is a clear clustering in two subtypes: 5 strains we characterize as A_1D_1 [GenBank (www.ncbi.nih) Accession Nos AF032432.1- AF032436.1] and 2 strains as A_1D_2 [GenBank (www.ncbi.nih) Accession Nos AF032430.1/31.1]. A_1 status is guaranteed by the SNP at Residue 75. D_1 status in the former 5 strains is clearly detected with 11 out of 12 homozygocities in the respective transition SNP positions. The transversion spot at *Residue* 319 is also A_1D_1-compatible. Only the Residue 30 presents a problem of a short, as it bears A-compatible homozygocity instead of AD heterozygocity. The intriguing thing is that the at the 2 A_1D_2 strains none of the 14 SNP positions expected to show heterozygocity presented any. They are homozygous for the D_2 –compatible alleles.

In detail, at 11 of the 12 biallelic SNPs (residues 60, 87, 117, 132, 135, 249, 252, 255, 258, 270, 279) plus the ones at Residues 75 and 319 the alleles called were A_1D_1 -compatible for all 5 A_1D_1-typed strains, but the allele at Residue 30 was not (homogenous A- compatible), nor the allele at Residue 228 (homogenous D2-compatible). The total was 5*(11+3)=65 out of 5*(12+3)=75 (86.66 % consistency rate) with steady positional distribution. For both the A_1D_2-typed strains, at all 12 of the 12 biallelic SNPs (residues 60, 87, 117, 132, 135, 249, 252, 255, 258, 270, 279) plus the ones at Residues 30, 75 and 319 the alleles called were homogenous D_2 -compatible giving a total of 2X15=30 out of 30 (100 % consistency rate). The grand total is 30+65=95 out of 30+75=105 (90.4% consistency rate). Only 5 incidents out of 105 examined positions (5.55%) are incompatible with the involved haplotypes and may point to specific reasoning. We cannot produce comparative SSCP-sequencing results for the Internet strains.

Discussion

In-houseAD strains

Our own three AD-serotype strains (Strain Nos 26, 35, 49) presented a A_1D_2 –compatible pattern. This was observed to some but not all of the originally 13 D_1-D_2 transition spots. The detected heterozygocities could be explained only if the presence of the D_2 haplotype is accepted. There were no other heterozygocities relative to homogenous types at other spots in any of these strains. Moreover, the heterozygocities were confirmed by visual check of the chromograms. Thus, we are induced to accept the non-random nature of the heterozygocity incidence at the exact positions of D_1-D_2 transitions. The only exception is the transition at Residue 69 (Fig. 2), which is unique to D_2 strain 23, implying either a very recent mutational event in D_2 (not yet passed to neither other D_2, nor the A_1D_x types) or a peculiarity of this particular strain.

Since the existence of an A strand (A_1, by force of the single SNP at Residue 75 – Fig. 2) is beyond any doubt by both RFLP and SSCP methods, we may assume that a clear and unilateral case of loss of heterozygocity (LOH) is in power, unidirectionally but not uniformly. There is no ground to deduce this is due either to some kind of imprint or any other mechanism. Though, it is obvious that the allele shifting occurs in favour of the D_2 alleles of the heterozygous positions. Moreover, this happens to different degrees (5 heterozygote positions out of 12 expected for strain 35; 7 out of 12 expected for strain 26; and 11 out of 12 for strain 49) which might imply that the process is still evolving. The difference in the distribution of these allele shifting events shows a discontinuous, possibly varied mechanism or chain of events which could not grade the 3 strains as different stages of a unique, clearly defined unidirectional process.
The 12th position (Residue 228 – Fig. 2) may be considered a real hot spot, showing no apparent consistency in its allelic status related to the subtype, due to reversal phenomena. However, this happens only in AD serotype, whereas in homogenous serotypes it shows absolute consistency. Alternatively, Residue 228 may be considered the very first position to shift to D_2 allelic condition.
The 5 GenBank A_1D_1 strains show at this position D_2-compatible homozygocity, which cannot be a LOH phenomenon, since the allele called (D_2) is not present, given that no sign of D_2 is detected at any of the other 11 biallelic informative loci in any of the 5 strains. It could be mutational allele shifting, perfectly explained by a Hot Spot status, occurring individually at the strain level and possibly followed by a LOH step. Or it might well be the first event towards the allele shifting. The A_1D_2 strains show either the expected heterozygocity (as is the case with our respective, in-house, strains) or the recurring D_2 -based homozygocity (GenBank retrieved), which in here could well be attributed to LOH causes.

Sequencing-SSCP

The comparative results of sequencing and SSCP, along with the observed allele shifting, lead to a series of questions regarding the discriminatory potential of SSCP. D_1 varies from D_2 in 12 recorded transition substitutions and a transversion substitution. As different subtypes have identical amplicon lengths, neither electrophoresis nor sequencing reveals D_1 or D_2 status lengthwise in heterozygotes. It is possible that changes at all 13 positions are needed to produce the homogenous D_2 SSCP type. We do not have a serotype D example with fused haplotype, (if a diploid D serotype exists in nature) to establish possibility of existence, subtype status and threshold and kinetics of the transformation. This is rather odd, because our 3 AD strains produced a clear A_1D_1 SSCP subtype, but transition-wise the haplotypes are partly A_1D_2 –compatible. Furthermore, each strain demonstrates a different degree of diversion from the A_1D_1 SSCP subtype.

Given that SSCP can demonstrably detect even one base substitution (**7**, **26)**, it is noteworthy that difference in 11 residues from the D_1 haplotype (e.g. strain 49) does not produce for the D strands of the AD strains a conformation diverging from D_1**.** The same goes for the LOH phenomenon, which alters the A_1 strands of all the A_1D_x strains in up to 7 residues (e.g strain 35) compared to the A_1 haplotype. SSCP's inability to detect up to 11 bases' substitutions may mean either a mechanism of reversal of conformation alterations, or a stochastic mechanism. As seven positions (strain 35, with the 7 LOHs) with 4 different bases per position create 4^7 possible conformations, it is obvious that observing identical conformations implies a factor which drastically reduces the possible -to- probable ratio and allows a much limited number of practically probable conformations. This, in turn, raises questions about the negative discriminatory power of SSCP (though not yet for the positive one): In essence, if two sequences do look different, they are, but if they look identical they might well not be. Of course the time parameter should be well taken into account, as the results suggest a dynamic genomics status, which may turn results from different points in time irrelevant. This is nicely observed in the case of strain 17, which was typically D_1 by both SSCP and sequencing, to shift fully to D_2 the second time it was sequenced. It was the only case we had sequencing data at two points in time, of which the earlier point in time was solidly related to SSCP data. The genomic fluidity suggested by these data disqualifies confirmation by successive repetitions of such time-sensitive experimental protocols.

Genbank retrieved AD sequences

The lack of any heterozygocities at the GenBank-retrieved AD strain sequences might lead to one of the following four proposals: The first possibility is the strains not being actually AD and to have been mistakenly serotyped, which is rather implausible due to the maturity of the techniques. The second possibility is that they are indeed AD but haploid, at least in this part of the genome (aneuploidy). The existence of D_2- compatible allele at Residue 228 in A_1/D_1-compatible haplotypes suggests the presence of a D_2 strand to impose the allele shifting (through LOH or otherwise).

The third possibility is that these strains are diploid but have reverted to a most stable condition, where LOH phenomena are completed (for the 2 A_1D_2 strains) or barely started (for the 5 A_1D_1 strains). It is not without importance that our own A_1D_x strains are highly heterozygotic, whereas the GenBank A_1D_x ones are homozygotic. The former show different levels of LOH, whereas the latter show complete LOH at all 15 informative loci. With the possible exception of technical allele- calling problems, a more stable genomic condition of the GenBank strains compared to our own is implied.
Such cohesiveness in status, displayed in 30 positions out of 30 at the 2 A_1D_2 GenBank strains may indicate the fact that the process of allele shifting towards the D_2 haplotype is complete in these 2 strains, and thus entails a visible and deterministic outcome, of unknown time determinant. It might be that this outcome is the loss of chromosomes towards aneuploidy or haploidy which might well explain the lack of heterozygocity. The non-causative relation between amplified sequence and the serotype would make such a fact plausible. To this end, it would be interesting if RFLP data (**25**) or other diploidy- sensitive data were made available. Though, the 5 A_1D_1 strains show uniformly an A-compatible allele at Residue 30. Since D_1 and D_2, at this position, show the same allele, it becomes clear that once the 2 A_1D_2 strains show the D-compatible allele, the D_1 strain is less "potent" to the A1 so as to suffer, not to cause, a LOH. Such a mechanism cannot be ruled out, but it is insufficiently backed by our current data and present knowledge. The problem is that according to Residue 228 the A_1D_1 strains start shifting to A_1D_2, but at the same time at Residue 30 the D_1 allele has shifted towards A_1. The ultimate A_1D_2 status needs a homozygous D-compatible allele at Residue 30, which makes 3 changes necessary at the same spot, which is not a very hot spot. Thus either Residue 228 is indeed a hot spot or the A-compatible homozygocity at Residue 30 is unexplained and dubious.
And the fourth possibility is the presence of a bug in the scoring of heterozygocities, either in the allele-calling or in the editing step. This would immediately put in doubt the credibility of the scoring system that typed the 5 strains as AD. Though, due to the extreme consistency in the results both horizontally (within each strain) and vertically (regarding the SNP positions) in these strains such extrapolation is rather unsubstantiated.
If heterozygocities do exist but the system fails to score them and scores forcedly a homozygocity, there must be a pattern at the ill-scored bases. A bug in base-calling at biallelic heterozygotic positions should normally create a 50%-odd consistency. The 7 AD strains with present a total of 15 positions each (highly informative residues 30, 75 and 319 and the twelve transitions) present a total of 105 informative positions. Assuming an A_1D_1 character for the former 5 strains and an A_1D_2 for the latter 2, we observe erroneous scoring at Residues 30 (A-homozygote instead of AD-heterozygote) and 228 (D_2 homozygote) for all the 5 A_1D_1 strains. We also observe 14 cases of erroneous scoring at each of the A_1D_2 strains, where D_2 homozygotes are scored instead of A_1D_2 heterozygotes. The total is 2*14+2*5=38 erroneous scoring incidents out of 7*15=105 positions (36.2%). Should a probabilistic mechanism be to blame, e.g. scoring of the allele in relative abundance in each heterozygocity as homozygote, there should be a more or less even distribution of allelic scoring in these positions or, at least, a 36.3% of erroneous results. This evenness should be at the grand total level and at each of the two degrees of freedom, i.e. vertically (per position) and horizontally (per strain). The examination vertically shows absolutely no anomalies and inconsistencies. Within each subtype, all positions were identically scored among the different strains. In the horizontal sense, the same was true for the A_1D_2 subtype, where all the 15 positions in doubt were scored identically, in accordance to the D_2 haplotype. For the A_1D_1 subtype there was a difference. One of the two misscored positions (Residue 30) was in accordance to the A_1 haplotype and the other (Residue 228) in accordance to D_2. Thus, in 15 positions there was one anomaly 1/15, a 6.66% which becomes 3.33% if the A_1D_2 (with no anomaly in haplotype) is taken into account. The fact that there were no inconsistencies (0% instead of the expected 50%) in the vertical sense should not be disregarded. From another perspective, the statistics were 5 anomalies to the relevant haplotypes (the Residue 228 in all 5 A_1D_1 strains) out of 38 erratic incidents, a mere 13.15%, which represent 5.55% at the grand total of 105 positions. This suggests a direct involvement of the haplotype as a factor in the creation of erratic results. Such data exclude any notion of probabilistic mechanism in the mistyping procedure and strongly advocate a stochastic phenomenon, which may be traced rather to the editing than to the allele-calling step. A bug in the post allele-calling editing step would be expected to create exactly such extreme consistencies, whether these apply to the correct alleles (in the case of the 5 A_1D_1 strains) or to the incorrect ones (in the case of the 2 A_1D_2 strains).

References

1. Βελεγράκη Α. και Καμπούρης Μ. Ε. (2003) «Μικροσυστοιχίες και Πολυπλεκτική Αλυσιδωτή Αντίδραση Πολυμεράσης: Επαναστατικές μέθοδοι Μοριακής Βιολογίας με εφαρμογές στην

Βιοϊατρική Πράξη» *Αρχεία Ελληνικής Ιατρικής, 20(4)*: 425-445.

2. Bennett JE, Kwon-Chung DJ Howard DH (1997), "Epidemiologic differences among serotypes of Cryptococcus neoformans" *Am J Epidemiol 105*, 582-6.

3. Buchanan KL, Murphy JW. 1998 "What makes Cryptococcus neoformans a pathogen?" *Emerg Infect Dis. 4(1):*71-83.

4. Canteros CE, Brudny M, Rodero L, Perrotta D, Davel G. (2002)["Distribution of Cryptococcus neoformans serotypes associated with human infections in Argentina"] *Rev Argent Microbiol. 34(4):*213-8.

5. Ellis DH, Pfeiffer TJ. (1990) "Ecology, life cycle, and infectious propagule of Cryptococcus neoformans." *Lancet. 13;336 (8720):*923-5.

6. Gaitanis G, Velegraki A, Alexopoulos EC, Chasapi V, Tsigonia A, Katsambas A. (2006) Distribution of Malassezia species in pityriasis versicolor and seborrhoeic dermatitis in Greece. Typing of the major pityriasis versicolor isolate M. globosa. *Br J Dermatol.;154:* 854-9.

7. Gathof BS, Sommer M, Podskarbi T, Reichardt J, Braun A, Gresser U, Shin YS. (1995) "Characterization of two stop codon mutations in the galactose-1-phosphate uridyltransferase gene of three male galactosemic patients with severe clinical manifestation". *Hum Genet. 96(6)* : 721-5.

8. Goldstein E, Rombo ON (1962) "Cryptococcal infection following steroid therapy" *Ann Intern Med 56:* 114

9. Ipsilantis I, Karpouzas DG et al (2008) "Effects of olive mill wastewaters on the community and function of arbuscular mycorrhizal fungi". *1st national panhellenic congress of the "Mikrobiokosmos" Society.*

10. Kambouris M. E., Reichard U., Legakis N. J. and Velegraki A. (1999) "Sequences from the aspergillopepsin *PEP* gene of *Aspergillus fumigatus*: Evidence on their use in selective PCR identification of Aspergillus species in infected clinical samples". *FEMS Immunol. Med. Microbiol. 25,* 255-64.

11. Καμπούρης Μ. Ε, Βελεγράκη Α., Λεγάκης Ν. Ι. (2000) "Διαφορική διάγνωση νηματοειδών μυκήτων με άπαξ περιοριστική πέψη προϊόντων αλυσιδωτής αντίδρασης πολυμεράσης". *Αρχεία Ελληνικής Ιατρικής 17:* 171-9.

12. Lazéra MS, Pires FD, Camillo-Coura L, Nishikawa MM, Bezerra CC, Trilles L, Wanke B. (1996) "Natural habitat of Cryptococcus neoformans var. neoformans in decaying wood forming hollows in living trees" *J Med Vet Mycol. 34(2):*127-31

13. Lazera MS, Salmito Cavalcanti MA, Londero AT, Trilles L, Nishikawa MM, Wanke B. (2000) "Possible primary ecological niche of Cryptococcus neoformans" *Med Mycol 38(5):*379-83.

14. Loftus BJ, Fung E, Roncaglia P. (2005) "The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans". *Science. 25;307(5713)*:1321-4.

15. Lynch JP 3rd, Schaberg DR, Kissner DG, Kauffman CA (1981)"Cryptococcus laurentii lung abscess" *Am Rev Respir Dis. 123(1):*135-8.

16. Mitchell TG, Perfect JR. (1995) "Cryptococcosis in the era of AIDS--100 years after the discovery of Cryptococcus neoformans." *Clin Microbiol Rev. 8(4):515-48*. Review.

17. Molez J (2001) "The historical question of AIDS in the 1960s in Cogo river basin area in relation to cryptococcal meningitis" , *Am J Trop Med Hyg 14:* 1401-7

18. Nosanchuk JD, Rudolph J, Rosas AL, Casadevall A. (1999) "Evidence that Cryptococcus

neoformans is melanized in pigeon excreta: implications for pathogenesis" *Infect Immun. 67(10):*5477-9

19. Perfect, J. R., and A. Casadevall. (2002) "Cryptococcosis" *Infect. Dis. Clin. N. Am. 16:*837-874.

20. Speed B & Dunt D. (1995) "Clinical and host differences between infections with the two varieties of Cryptococcus neoformans." *Clin Infect Dis. 21(1)*:28-6.

21. Sorrell TC, Brownlee AG, Ruma P, Malik R, Pfeiffer TJ, Ellis DH.(1996) "Natural environmental sources of Cryptococcus neoformans var. gattii." *J Clin Microbiol. 34(5):*1261-3.

22. Tanaka KI, Miyazaki T et al. (1996) "Detection of Cryptococcus neoformans gene in patience with pulmonary cryptococcosis" *J Clin Microbiol 34:* 2826-8

23. Velegraki A., Kambouris M. E. et al (1999). "Rapid extraction of fungal DNA from clinical samples for PCR amplification". *Medical Mycol. 37,* 69-73.

24. Velegraki A., Kambouris M. E. et al. (1999) "Identification of medically significant fungal genera by Polymerase Chain Reaction followed by Restriction Enzyme Analysis". *FEMS Immunol. Med. Microbiol. 23,* 303-12.

25. Velegraki A, Kiosses VG et al. (2001) "Prospective use of RFLP analysis on amplified Cryptococcus neoformans URA5 gene sequences for rapid identification of varieties and serotypes in clinical samples.". *Med Mycol 39*: 409-17.

26. Velegraki A, Menounos PG et al (2002). "Typing Greek Clinical isolates of C. neoformans var neoformans by PCR-SSCP". *5th International Conference Cryptococcus & cryptococcosis, Adelaide, Australia.*

27. Zimmerman LE, Rappaport H. (1954)" Occurrence of Cryptococcosis in patients with malignant disease of reticuloendothelial system." *Am J Clin Pathol. 24(9):*1050-72

28. Zuger A, Louie E, Holzman RS, Simberkoff MS, Rahal JJ. (1986) "Cryptococcal disease in patients with the acquired immunodeficiency syndrome. Diagnostic features and outcome of treatment." *Ann Intern Med. 104(2):*234-40

Αλληλουχικό υπόβαθρο της οροτυποποιητικής και υποτυποποιητικής αποτελεσματικότητας της PCR με εναρκτές URA-A/B στον *Cryptococcus neoformans* και ενδείξεις για μονόδρομη πολλαπλή αλληλική εκτροπή

Τούκας Δημήτριος* (1), Καμπούρης Μανούσος Εμμ *.(2), Γαϊτάνης Γεώργιος (1,3), Μενούνος Παναγιώτης (4) και Βελεγράκη Αριστέα (1)

(1) Τμήμα Ιατρικών Εργαστηρίων, Σχολή Επαγγελμάτων Υγείας-Πρόνοιας, ΑΤΕΙ Αθηνών, Αγ. Σπυρίδωνος, Αιγάλεω 12210.
(2) Ειδικό Εργαστήριο Μυκητολογίας, Ιατρική Σχολή, ΕΚΠΑ, Μικράς Ασίας 75-77 Γουδή, 11527
(3) Κλινική Δερματικών και Αφροδισίων Νοσημάτων, Ιατρική Σχολή, Πανεπιστήμιο Ιωαννίνων, Λεωφόρος Σ. Νιάρχου, Ιωάννινα 45110
(4) Εργαστήριο Μοριακής Βιολογίας, Σχολή Αξιωματικών Αδελφών Νοσοκόμων, Αθήνα

* Συγγραφείς ταυτόσημης συμμετοχής

Περίληψη

Το γενετικό υπόβαθρο της μοριακής υποτυποποίησης σε στελέχη *Cryptococcus neoformans* που επιτυγχάνεται με μεθοδολογία PCR-RFLP ή PCR-SSCP οδήγησε, μέσω αλληλουχοποίησης και στοίχισης αλληλουχιών στην ανίχνευση Μονονουκλεοτιδικών Πολυμορφισμών (SNP) με σταθερό αλληλικό καθεστώς και σαφή ανυσματική αλληλική εκτροπή. Εντός και μεταξύ των καθιερωμένων υποειδικών ταξινομικών βαθμίδων (ποικιλία και ορότυπος). Ο αλληλικός τύπος των κατά σύμβαση αλληλουχιών από οροτυποποιημένα και υποτυποποιημένα στελέχη υποδεικνύει την πιθανότητα οροτυπικής υποστροφής δια βαθμιαίας αλληλικής εκτροπής. Η φορά της υποστροφής είναι από τον ορότυπο Α μέσω του AD στον D, αν και παραμένει άγνωστο το αν αυτά τα βήματα μπορούν να γίνουν διαδοχικά από το ίδιο στέλεχος..

Λέξεις -Κλειδιά
Cryptococcus neoformans, γονίδιο URA5, PCR, SSCP, Ορότυπος AD, Μονονουκλεοτιδικός Πολυμορφισμός (SNP), Απώλεια Ετεροζυγωτίας (LOH)