

1
2
3 **Genotyping *Candida albicans* from *Candida* leukoplakia and non-**
4 ***Candida* leukoplakia shows no enrichment of multilocus sequence typing**
5 **clades but enrichment of ABC genotype C in *Candida* leukoplakia**
6

7
8 **Mohammed H. Abdulrahim,^{1#} Brenda A. McManus,^{2#} Stephen R. Flint,¹ and David**
9 **C. Coleman^{2*}**

10
11
12 *Division of Maxillofacial Surgery, Oral Medicine and Oral Pathology, Dublin Dental*
13 *University Hospital, University of Dublin, Dublin 2, Republic of Ireland¹ Microbiology*
14 *Research Unit, Division of Oral Biosciences, Dublin Dental University Hospital, Trinity*
15 *College Dublin, University of Dublin, Dublin 2, Republic of Ireland,²*

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 *Corresponding author: Mailing address: Microbiology Research Unit, Division of Oral Biosciences, Dublin
32 Dental University Hospital, University of Dublin, Trinity College Dublin, Dublin 2, Republic of Ireland.
33 Phone: +353 1 6127276. Fax: +353 1 6127295. E-mail : david.coleman@dental.tcd.ie.

34 # These authors contributed equally to this work.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

ABSTRACT

Oral leukoplakias are histopathologically-diagnosed as *Candida* leukoplakia or non-*Candida* leukoplakia by the presence or absence of hyphae in the superficial epithelium. *Candida* leukoplakia lesions have significantly increased malignant potential. *Candida albicans* is the most prevalent fungal species associated with oral leukoplakia and may contribute to malignant transformation of *Candida* leukoplakia. To date, no detailed population analysis of *C. albicans* isolates from oral leukoplakia patients has been undertaken. This study investigated whether specific *C. albicans* genotypes were associated with *Candida* leukoplakia and non-*Candida* leukoplakia in a cohort of Irish patients.

Patients with histopathologically-defined *Candida* leukoplakia (n=31) or non-*Candida* leukoplakia (n=47) were screened for *Candida* species by culture of oral rinse and lesional swab samples. Selected *C. albicans* isolates from *Candida* leukoplakia patients (n=25), non-*Candida* leukoplakia patients (n=19) and oral carriage isolates from age and sex matched healthy subjects without leukoplakia (n=34) were subjected to multilocus sequence typing (MLST) and ABC genotyping.

MLST revealed that the clade distribution of *C. albicans* from both *Candida* leukoplakia and non-*Candida* leukoplakia lesions overlapped with the corresponding clade distributions of oral carriage isolates and global reference isolates from the MLST database indicating no enrichment of leukoplakia-associated clones.

Oral leukoplakia isolates were significantly enriched with ABC genotype C (12/44, 27.3%), particularly *Candida* leukoplakia isolates (9/25, 36%), relative to oral carriage isolates (3/34, 8.8%). Genotype C oral leukoplakia isolates were distributed in MLST clades 1,3,4,5,8,9 and 15, whereas genotype C oral carriage isolates were distributed in MLST clades 4 and 11.

1 **INTRODUCTION**

2 Several *Candida* species cause opportunistic infections in humans, the most
3 prevalent and pathogenic of which is *Candida albicans*. *Candida albicans* is also the most
4 common *Candida* species isolated from the oral cavity as both a commensal and
5 pathogenic organism in healthy individuals and those with underlying disease [1].

6 Oral leukoplakia is used as a general, clinical descriptive term to describe oral
7 white patches or lesions of questionable risk, having excluded other known diseases that
8 carry no increased risk for cancer [2]. These lesions can be further defined based on
9 histopathological diagnosis or clinical history. Lehner [3] introduced the term “*Candida*
10 leukoplakia” to describe chronic oral *Candida* infection presenting in the form of
11 leukoplakia, for which increased malignant potential has been reported [4] compared to
12 simple leukoplakia. *Candida* leukoplakia usually presents clinically as a well demarcated,
13 rough, raised, white plaque-like lesion that cannot be rubbed off. The lesion may have an
14 homogenous white surface, (homogenous leukoplakia), or in contrast a non-homogenous,
15 nodular or speckled appearance, with an erythematous base. The commissure is the most
16 common site affected, but it may also affect the palate and tongue. These lesions may be
17 associated with angular cheilitis. *Candida albicans* is by far the most prevalent species
18 associated with oral leukoplakia lesions [5].

19 *Candida* leukoplakia (CL) lesions are difficult to distinguish from non-*Candida*
20 leukoplakias (NCLs) clinically, but the presence of invading *Candida* hyphae in the
21 superficial layer of epithelium accompanied by infiltration of polymorphic neutrophils
22 histologically distinguish CL lesions [4,6]. The role of *Candida* in inducing keratotic
23 changes and cellular atypia is uncertain, and there is some controversy regarding the
24 initiation of epithelial hyperplasia by *Candida*. Jepsen and Winther [7] suggested that
25 *Candida* invades a pre-existing hyperplastic lesion, whereas in contrast Cawson and

1 Lehner [4] proposed that *Candida* infection is the primary cause of CL. Holmstrup and
2 Besserman [8] demonstrated reversion of non-homogenous CL to the homogenous type
3 after the use of topical antifungal agents. McCullough et al. [9] reported a strong
4 statistical association between increased oral yeast density, oral epithelial dysplasia and
5 oral squamous cell carcinoma, with the degree of epithelial dysplasia correlating with
6 increased oral yeast density. Based on these findings, the authors hypothesised that the
7 progression of *Candida* leukoplakia to dysplasia is promoted by *C. albicans*. With NCL
8 lesions, the inference is that *Candida* species isolated from these lesions are non-
9 pathogenic, commensal yeast blastospore contaminants.

10 Detailed molecular epidemiological and population studies based on *C. albicans*
11 isolates recovered from CL lesions are, to date, mostly lacking. This is most likely due to
12 practical difficulties in obtaining sufficient isolates from such lesions for detailed studies
13 due to the relative rarity of the condition and the requirement for biopsy and
14 histopathological investigation. Furthermore, histopathological diagnosis of such lesions
15 has previously been quite vague. A previous report based on biotyping experiments
16 suggested that *C. albicans* isolates recovered from oral leukoplakia lesions in 17 separate
17 patients may be genetically different to those which typically colonise the normal oral
18 mucosa [5]. In contrast, another study that used PCR-based fingerprinting methods using
19 two interrepeat primer combinations and a minisatellite-specific primer failed to find
20 clonal restrictions among 38 *C. albicans* strains recovered from 17 patients with CL [10].
21 However, in this study not all lesions were histopathologically confirmed as CL.

22 Multilocus sequence typing (MLST) of *C. albicans* has been used widely to assess
23 the genetic relatedness of isolates recovered from disparate geographic locations and for
24 *C. albicans* population analysis [11,12,13]. MLST is a DNA-based method that examines
25 nucleotide sequence variation in seven housekeeping genes and is highly reproducible

1 between different laboratories. A curated, internet-based *C. albicans* MLST database
2 (<http://calbicans.mlst.net>) has facilitated the comparison of isolates from a wide variety of
3 anatomical sites and geographic locations. *Candida albicans* isolates can also be ABC
4 genotyped based on the presence or absence of an intron in 25S rDNA [14]. Although this
5 method has previously been described as a helpful confirmatory test in instances where
6 isolates differ by one or more single nucleotide polymorphisms (SNPs) in MLST [15],
7 and has been used to demonstrate geographical and temporal differences amongst *C.*
8 *albicans* isolates [16], the method examines only one genetic marker, has a low
9 discriminatory power (i.e. discriminates isolates into one of three genotypes) and
10 therefore should not be used as a definitive test for isolate relatedness [15]. Genotypes A
11 and B predominate amongst *C. albicans*, whereas genotype C is quite rare [17].

12 The purpose of the present study was to characterise the population of *C. albicans*
13 isolates recovered from histopathologically-defined CL and NCL lesions from a cohort of
14 Irish oral leukoplakia patients using MLST and ABC genotyping in order to determine if
15 these lesions could be associated with specific *C. albicans* lineages.

16

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

MATERIALS AND METHODS

Ethics statement. Ethical approval for this study was obtained from the St. James's Hospital (SJH) and Adelaide and Meath Hospitals including the National Children's Hospital (AMNCH) Research Ethics Committee, Dublin, Ireland. Prior to enrollment in the study, all participants were provided with comprehensive patient information documentation and all participants included in the study provided written consent. All documentation provided to patients, including consent forms were pre-approved by the Research Ethics Committee. This paper does not include any identifying, or potentially identifying, patient information.

Study group. Oral rinse and swab samples were collected from 78 patients with oral leukoplakia attending the Oral Medicine and Dysplasia clinics at the Dublin Dental University Hospital (DDUH) between December 2006 and March 2009 (Table 1). Patients were examined and followed-up by two oral medicine consultants and an oral maxillofacial consultant. A full medical and dental history was recorded and a comprehensive oral examination undertaken. A clinical diagnosis of oral leukoplakia was made as described previously [2]. A histopathological diagnosis of CL was made based on the detection of hyphae and hyphal-associated polymorphonuclear leukocytes in CL biopsies. Similarly, the absence of hyphae in leukoplakia biopsies was used to define NCL lesions. Squamous cell carcinoma developed in two CL patients at the lesional site within the follow up two-year period. None of the NCL group developed malignant change to date. The mean age of the patient group was 57.8 years (range 29-87 years) and the sex distribution was 40 (51%) male and 38 (49%) female. Forty-nine patients were smokers (62.8%), 20 of whom were diagnosed with CL. Fifteen (48.3%) of the CL patients and 11 (23.4%) NCL patients wore upper dentures.

1 Oral rinse samples were also taken from a control group of 110 healthy volunteers
2 recruited from the DDUH Accident and Emergency Department for comparison. Of these,
3 56 (50.9%) were male and 54 (49.1%) were female, the mean age was 54.1 years (range
4 29-78 years) and none had oral leukoplakia. These samples were taken during the same
5 time period as the leukoplakia samples (i.e. December 2006-March 2009). Twenty-one
6 healthy volunteers were smokers (19.0%) and six (5.5%) wore upper dentures.

7 Patients with leukoplakia and healthy volunteers were excluded from the study if
8 they met any of the following criteria: pregnancy or lactation, diabetes or asthma, steroid
9 treatment during the last year, antibiotic or anti-fungal treatment in the previous six
10 months.

11 **Histopathological diagnosis of CL and NCL lesions.** A DDUH oral surgeon
12 performed incisional lesional biopsies. Biopsy samples were fixed in 10% (v/v) neutral
13 buffered formalin (4% (v/v) formaldehyde). Histopathological investigation of biopsy
14 samples was performed at the Central Pathology Laboratory at St. James's Hospital,
15 Dublin. Fixed tissues were dehydrated with 70-100% (v/v) ethanol, cleared with xylene
16 and embedded in paraffin wax. Tissue samples were then cut into 6-8 μm thick sections
17 and routinely stained with haematoxylin and eosin. As candidal hyphae are poorly stained
18 by hematoxylin/eosin, the periodic acid-Schiff (PAS) stain was also used, as hyphal
19 structures stain strongly with PAS [18]. An oral and maxillofacial pathologist made
20 diagnoses based on the histological features of each specimen.

21 **Sampling of the oral cavity and leukoplakia lesions.** Oral rinse samples were
22 performed as described previously [19]. *Candida* isolates were recovered by culture
23 following incubation at 37°C for 48 h on the chromogenic culture medium CHROMagar
24 *Candida*TM (CHROMagar Company, Paris, France). This medium permits the

1 presumptive identification of several clinically important *Candida* species including *C.*
2 *albicans* based on colony colour and morphology [20].

3 Swab samples were taken using sterile cotton transport swabs and then transferred
4 to alginate gel transport medium (Venturi Transystem, Copan Italia s.p.a, Brescia, Italy).
5 Swab samples were taken by rubbing the entire surfaces of the lesion for 30 s; the
6 swabbed areas sampled depended on the size of the lesion. Swab samples were plated
7 within 2 h on to CHROMagar *Candida*TM medium and incubated as above.

8 **Identification of *Candida* isolates.** Following incubation, CHROMagar
9 *Candida*TM plates were examined, and the relative abundance of each colony type present
10 was recorded. Selected examples of each colony type were chosen for detailed analysis.
11 Isolates were initially presumptively identified on the basis of colony color and
12 morphology [20,21] and were definitively identified by substrate assimilation profiles
13 using the API ID 32C yeast identification system (BioMérieux, Marcy l’Etoile, France)
14 [22].

15 **DNA extraction.** Genomic DNA from *C. albicans* isolates was extracted using the
16 Qiagen DNeasy blood and tissue kit (Qiagen Crawley, West Sussex, UK) according to the
17 manufacturer’s instructions.

18 **ABC genotyping.** Template DNAs extracted from isolates were assigned to
19 genotypes A, B or C based on the differential PCR amplification of the 25S rRNA gene,
20 as previously described [14].

21 **MLST.** Twenty-five *C. albicans* isolates from CL lesions and 19 isolates from
22 NCL lesions, from separate patients in each case (with the exception of two isolates,
23 CL12 and CL122 which were recovered from separate CL lesions in patient 12), were
24 subjected to MLST as described previously [12]. For comparison, MLST was also
25 performed on 34 oral carriage (OC) *C. albicans* isolates recovered from oral rinse samples

1 of 34 separate control healthy subjects attending the DDUH Accident and Emergency
2 Department. All DNA sequencing reactions were performed commercially by Source
3 Bioscience LifeSciences (Dublin, Ireland) using ABI 3730xl DNA analysers and dye-
4 labelled terminators (Applied Biosystems, Foster City, CA, USA). Sequences were
5 analysed using BioNumerics version 5.0 software (Applied Maths NV, Saint-Martens-
6 Latem, Belgium) and the online *C. albicans* MLST database (<http://calbicans.mlst.net>).
7 All MLST data generated in the present study have been deposited in the *C. albicans*
8 MLST database (<http://calbicans.mlst.net>).

9 The genetic relatedness of the isolates investigated in the present study to each
10 other and to selected strains from the online MLST database was evaluated using the
11 software program MEGA version 5 to generate dendrograms based on the unweighted-
12 pair group method with arithmetic averages (UPGMA) [23]. The UPGMA clustering
13 algorithm defines MLST clades that correlate well with those previously defined by DNA
14 fingerprinting using the *C. albicans*-specific repetitive sequence-containing Ca3 probe
15 [24].

16 The online Based Upon Related (eBURST) algorithm (<http://eburst.mlst.net>) was
17 used to subdivide MLST allelic profile datasets into hypothetical non-overlapping groups
18 of clonal complexes (CCs) composed of a single putative founder diploid sequence type
19 (DST) and its closely related descendant DSTs [25,26].

20 **Statistical Analyses.** Statistical analyses such as Fisher's exact tests and two-
21 tailed Student's *t*-tests for independent sample proportions were carried out using
22 Graphpad software (Graphpad Software Inc., CA, USA). A power analysis was
23 undertaken based on the comparison of two independent samples. The parameters for this
24 one sided test used a desired power value of 75% with an α value 0.05
25 (<http://stat.ubc.ca/~rollin/stats/ssize/b2.html>).

26

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25**RESULTS****Oral leukoplakia and healthy control subjects.**

Clinical details of the oral leukoplakia patients are shown in Table 1. Seventy-eight patients with oral leukoplakia were enrolled in the study, of whom 31/78 (39.7%) had histopathologically-defined CL. Biopsies from the remaining 47 patients (60.3%) were classified as NCL following histopathological investigation (Table 1). None of the healthy subjects included in the study had clinical signs of oral leukoplakia. The buccal mucosa was the most prevalent lesional site in the CL group (22/31; 71%), followed by the tongue (4/31; 12.9%), palate (3/31; 9.7%), gingivae (1/31; 3.2%) and alveolar ridge (1/31; 3.2%) (Table 1). The buccal mucosa was also the most prevalent lesional site in the NCL group (21/47; 44.7%), followed by the by the tongue (9/47; 19.1%), alveolar ridge (6/47; 12.8%), palate (5/47; 10.6%), the floor of the mouth (4/47; 8.5%) and gingivae (2/47; 4.3%) (Table 1).

Candida detection by oral rinse sampling.

Candida isolates were recovered from oral rinse samples in 55/78 oral leukoplakia patients (70.5%). *Candida albicans* was the most prevalent *Candida* species recovered and was isolated from all 55 *Candida*-positive oral leukoplakia patients, either alone (48 patients) or in combination with other *Candida* species or *Saccharomyces cerevisiae* (seven patients) (Table 1). Oral rinse sampling of 24/31 (77.4%) patients with CL yielded *Candida* species, all of whom harboured *C. albicans*, either alone (20 patients) or in combination with other *Candida* species or *S. cerevisiae* (four patients) (Table 1). Oral rinse sampling of 31/47 (66%) NCL patients yielded *Candida* species, all of whom harboured *C. albicans*, either alone (28 patients) or in combination with other *Candida* species (3 patients) (Table 1).

The average *Candida* cell density recovered by oral rinse sampling of the 55 *Candida*-positive patients was 719 cfu/ml (range 2 cfu/ml to approximately 6000 cfu/ml) (Table 1). Oral rinse sampling of 34/110 (30.9%) healthy control subjects yielded *C.*

1 *albicans*. The median and mean average *Candida* cell density values recovered from these
2 34 *Candida*-positive controls was 18 and 32 cfu/ml, respectively (range 1-150 cfu/ml). No
3 other yeast species were recovered from the healthy control subjects.

4 ***Candida* detection by lesional swab sampling.** *Candida* species were recovered
5 by swab sampling of oral leukoplakia lesions from 50/78 (64.1%) of the total oral
6 leukoplakia patient cohort. *Candida albicans* was the predominant species recovered and
7 was isolated from 48/50 (96%) *Candida*-positive patients either alone (42 patients) or in
8 combination with other *Candida* species or *S. cerevisiae* (six patients). One patient
9 (patient 49) yielded *C. krusei* only and another patient (patient 36) yielded *C. glabrata*
10 only using the swab method. Both of these patients yielded *C. albicans* by oral rinse
11 (Table 1).

12 Swab sampling of lesional tissues from 27/31 (87.1%) patients with CL yielded
13 positive cultures for *Candida* species, all of which consisted of *C. albicans* either alone
14 (23 patients) or in combination with other *Candida* species or *S. cerevisiae* (Table 1).
15 Swab sampling of lesional tissues from 23/47 (48.9%) patients with NCL yielded positive
16 cultures for *Candida* species, 21/23 (91.3%) of which consisted of *C. albicans* either
17 alone (20 patients) or in combination with *C. tropicalis* (patient 67). One patient yielded
18 *C. glabrata* only (patient 36) and one patient yielded *C. krusei* only (patient 49) (Table 1).

19 Interestingly, a comparison of oral rinse and lesional swab *Candida* culture data
20 revealed that swab samples from oral lesions in eight patients [five CL patients (03, 07,
21 08, 09 and 28) and three NCL patients (34, 48 and 56)] were culture-positive for *Candida*,
22 whereas the corresponding oral rinse samples were culture-negative (Table 1). The
23 opposite was true for patient 40, from whom *Candida* was recovered by oral rinse, but not
24 by lesional swabbing (Table 1).

25

1 **MLST analysis of *C. albicans*.** Isolates from CL (n=25), NCL (n=19) and healthy
2 controls (n=34), were analysed with the consensus *C. albicans* MLST scheme [12].
3 MLST resulted in a dataset of 2,883 bp for each isolate. Single nucleotide polymorphisms
4 (SNPs) were identified in 84 of the 2,883 bp (2.9%) analysed. The *AAT1a* gene yielded 11
5 SNPs, (13.1%) the *ACCI* gene yielded six (7.1%) SNPs, the *ADPI* gene yielded 14
6 (16.7%) SNPs, the *MPIb* gene yielded 12 (14.2%) SNPs, the *SYA1* gene also yielded 11
7 (13.1%) SNPs, the *VPS13* gene yielded 15 (17.9%) SNPs and the *ZWF1* gene yielded 15
8 (17.9%) SNPs.

9 New allelic profiles and DSTs were identified in 56 isolates (Table 2). Of the
10 seven loci investigated for each isolate, five new alleles were identified including four
11 new *ZWF1b* alleles, and one *MPIb* allele. Only two DSTs were identified in duplicate;
12 The newly identified DST2108 was identified in an isolate (OL37) recovered from an
13 NCL lesion and in an isolate (OC304) recovered from a healthy individual. Another
14 newly identified DST, 2105, was identified in isolates CL12 and CL122, however, these
15 isolates were recovered from separate CL lesions in the same patient, patient 12 (Table 2
16 and Figure 1).

17 A UPGMA dendrogram was constructed using the MLST allelic profile data
18 resulting from 44 isolates recovered from oral leukoplakia lesions, 34 OC isolates and 42
19 isolates representative of previously documented MLST clades previously included in the
20 *C. albicans* MLST database (data not shown), to determine the genetic relatedness of
21 isolates in the present study to those previously included in the MLST database, and to
22 assign newly identified DSTs to clades. This tree identified the presence of distinct MLST
23 clades as previously described [17]. A smaller UPGMA dendrogram was constructed
24 based on the concatenated MLST SNPs of only the 78 isolates included in the current
25 study (Figure 1). This dendrogram shows that isolates recovered from both CL and NCL

1 patient groups are distributed amongst 11 distinct MLST clades, and OC isolates are
2 distributed amongst **nine** distinct MLST clades, with clade 1 predominant in all three
3 groups of patients (Figure 1 and Table 2). Of the 78 isolates investigated by MLST, 27/78
4 (34.6%) belonged to *C. albicans* MLST clade 1, consisting of nine CL isolates, eight NCL
5 isolates and 10 OC isolates. A further 12/78 (15.4%) isolates belonged to MLST clade 2,
6 of which nine were OC isolates and three were NCL isolates. There was an absence of CL
7 isolates in clade 2. The third most predominant clade was clade 4, which contained 9/78
8 (11.5%) isolates including six OC isolates, two CL isolates and one NCL isolate.
9 Interestingly, all four isolates belonging to clade 3 in the present study were recovered
10 from CL lesions, and all three isolates belonging to clade 9 in the present study were
11 recovered from NCL lesions (Figure 1 and Table 2).

12 The level of heterozygosity amongst the 84 polymorphic nucleotides were
13 determined for the *C. albicans* isolates investigated in the present study, as significant
14 losses of heterozygosity can be associated with minor genetic switches or microvariation
15 [15,27]. The average number of heterozygous sites in isolates recovered from CL lesions
16 was 16.9 ± 6.2 sites compared to 13.9 ± 14.1 sites in isolates recovered from NCL lesions
17 and 13.4 ± 5.4 sites in the OC isolates recovered from healthy patients (Table 2). These
18 differences are most likely the result of the minor differences in clade prevalence amongst
19 isolates from each patient group.

20 According to eBURST analysis, an alternative method of predicting evolutionary
21 patterns and founding genotypes, the most predominant clonal complex (CC) identified
22 amongst 21/78 (26.9%) isolates from the present study was CC1. CC1 isolates were
23 recovered from seven CL patients, seven NCL patients and seven OC subjects (Table 2).
24 Five isolates belonged to CC2 and six to CC3 (Table 2). Singletons that were not assigned

1 to any CCs were identified in isolates recovered from 8/25 (32%) CL patients, 6/19
2 (31.6%) NCL patients and 12/34 (35.3%) OC isolates (Table 2).

3 **ABC genotyping of *C. albicans* isolates.** The same 78 isolates that were
4 investigated by MLST were subjected to ABC genotyping, including isolates from CL
5 (n=25), NCL (n=19), and OC (n=34). Genotype A was the most common genotype and
6 was identified in 51/78 (65.4%) isolates. Twelve isolates (15.4%) were identified as
7 genotype B and 15 (19.2%) were identified as genotype C (Table 2). Of the 34 OC
8 isolates, 26 (76.5%) were identified as genotype A, five were identified as genotype B
9 (14.7%), and the remaining three were genotype C (8.8%). In contrast, only 25/44
10 (56.8%) oral leukoplakia isolates (12 CL and 13 NCL) belonged to genotype A (Table 3).
11 Genotype B was identified in seven (15.9%) oral leukoplakia isolates (four CL and three
12 NCL). Interestingly, 12/44 (27.3%) oral leukoplakia isolates were identified as genotype
13 C, (nine CL and three NCL). The proportions of genotypes A-C among CL and OC
14 isolates were significantly different ($P < 0.02$ with Fisher's exact test). Isolates from CL
15 lesions were significantly enriched with genotype C 9/25 (36%) when compared to OC
16 isolates (3/34, 8.8%) from healthy carriers ($P < 0.05$ with two sample Student's *t* test). Due
17 to the small number of CL isolates available for investigation in the present study, a
18 power analysis was undertaken to determine if the sample size was sufficient to reliably
19 detect this increased prevalence of genotype C. Using a desired power value of 75% with
20 an α value 0.05, a sample size of 25 was indicated for each sample group. As 25 CL
21 isolates and 34 OC isolates were included in the study, the results of the power analysis
22 indicated that sufficient isolates were investigated to support the significance of the
23 observed increased prevalence of genotype C isolates amongst CL isolates relative to OC
24 isolates.

25

26

DISCUSSION

1
2 The present study observed that *Candida* isolates were more frequently recovered
3 by oral rinse sampling of patients with oral leukoplakia (70.5% positive) than of healthy
4 controls (30.9% positive). Unsurprisingly, *C. albicans* was the most frequently identified
5 *Candida* species, and was recovered from all *Candida*-positive oral leukoplakia patients
6 either alone, or in combination with other species. No *Candida* species other than *C.*
7 *albicans* were recovered from the healthy control population. Oral leukoplakia patients
8 were histopathologically-diagnosed as having CL or NCL in the present study. *Candida*
9 species were recovered from swab sampling of 64.1% of the oral leukoplakia patients, and
10 in 87.1% of patients with CL. Interestingly, comparative oral rinse and lesional swab
11 sampling methods revealed that in eight oral leukoplakia patients, lesional swab yields of
12 *Candida* species did not correlate with corresponding oral rinse sampling *Candida* yields
13 (Table 1). It is possible that the oral rinse sampling method, and to a lesser extent swab
14 sampling, was not vigorous enough to dislodge strongly adherent and/or invasive *Candida*
15 blastospores or hyphae from these CL lesions.

16 MLST is a very effective tool for *C. albicans* population analyses [11]. Overall,
17 the population analysis based on MLST suggests that no clonal enrichment exists in
18 isolates recovered from CL or NCL lesions. The isolates recovered from both CL and
19 NCL patient groups are distributed amongst 11 distinct MLST clades, and OC isolates are
20 distributed amongst nine distinct MLST clades. Clade 1 predominated in all three groups
21 of patients with 27 isolates (34.6%) belonged to this clade. Twelve isolates (15.4%)
22 belonged to MLST clade 2, and nine (11.5%) isolates belonged to MLST clade 4. This
23 study therefore supports previous findings [17,28], that clades 1, 2 and 4 comprise more
24 than the 2/3 of isolates in the MLST database.

1 Interestingly, there was an absence of CL isolates in clade 2, whereas 9/12 (75%)
2 clade 2 isolates consisted of OC isolates representing 26.5% (9/34) of all OC isolates
3 included in the study. All four isolates belonging to clade 3 in the present study were
4 recovered from CL lesions. Furthermore, all three isolates belonging to clade 9 in the
5 present study were recovered from NCL lesions. It is likely that increasing the numbers of
6 CL, NCL and OC isolates in larger future studies would result in the diversification of
7 these clades (Figure 1).

8 Currently, there are 2,218 *C. albicans* isolate profiles in the MLST database
9 (<http://calbicans.mlst.net/>) comprising 2,086 unique allelic profiles and their
10 corresponding DSTs (date accessed: 10th April 2013), of which 373 (17.9%) isolates are
11 described as oral isolates comprising 290 distinct DSTs. According to the eBURST
12 algorithm, these oral isolate DSTs can be divided into 24 different CCs, with CC1
13 predominating. The remaining 131/290 (45.2%) DSTs are classified as singletons. In the
14 present study, analysis of 78 *C. albicans* MLST profiles using the eBURST algorithm
15 showed that 50 isolates were distributed among 17 CCs, with CC1 predominating.
16 Twenty-six (33.3%) isolates were assigned as singletons. This suggests that, like the
17 UPGMA clade-based method, eBURST identifies no significant clonal enrichment in
18 isolates recovered from oral leukoplakia patients. Overall, our MLST data correlates with
19 previous biotyping studies [10] which failed to find clonal distinction among *C. albicans*
20 strains recovered from CL lesions.

21 ABC genotyping of *C. albicans* isolates was included in the present study to
22 provide further isolate discriminatory data. Similarly to previous studies [13,16,17,29]
23 genotype A predominated in all groups examined, the majority (92.6%) of isolates
24 belonging to MLST clade 1 were identified as genotype A. Surprisingly, genotype C was
25 identified more frequently in the current study than in the aforementioned studies,

1 particularly in isolates from oral leukoplakia lesions (Table 3), being identified in 9/25
2 (36%) and 3/19 (15.8%) of isolates recovered from CL and NCL lesions, respectively
3 (Table 3). In contrast, only three genotype C isolates were identified among the 34 (8.8%)
4 OC isolates investigated. However, the MLST data clearly did not support any suggestion
5 of clonal enrichment (Figure 1 and Table 2). It is important to emphasise that MLST has
6 been shown to be an exquisitely informative tool for inferring clonal relationships
7 between *C. albicans* isolates [17] in contrast to ABC genotyping, which examines only
8 one locus and can only separate isolates into three groups. ABC genotypes A and B are
9 determined by the absence or presence of an intron in the 25S rRNA gene, respectively.
10 Genotype C is a hybrid of the other two genotypes, containing the intron in only one
11 allele. On balance, it is unlikely that the observed genotype C enrichment among isolates
12 from leukoplakia lesions reflects a specific clonal enrichment. This suggestion is
13 supported by the findings that ABC genotype C was identified in *Candida* leukoplakia
14 isolates belonging to MLST clades 1, 3, 4, 8 & 15, and in non-*Candida* leukoplakia
15 isolates belonging to MLST clades 1, 5 & 9.

16 The oral leukoplakia isolates investigated in the present study might represent a
17 geographical clonal population of *C. albicans* from Ireland. However, this seems unlikely
18 as the OC control isolates were also from the same geographical location but exhibited a
19 different ABC genotype distribution (Table 3). Further investigations based on a larger
20 patient numbers are required and should include individuals from disparate geographical
21 locations and ethnic backgrounds in order to decipher this phenomenon further.
22 Additional investigations might involve comparative SNP microarray or whole genome
23 sequence analysis of several CL and NCL isolates from each ABC genotype.

24 Very few previous studies actually correlated *C. albicans* ABC genotypes with
25 specific medical conditions. Most previous studies that used ABC genotyping did so as a

1 secondary typing method in combination with another, more discriminatory typing
2 method such as microsatellite typing or MLST. *Candida albicans* ABC genotypes tend to
3 be more often associated with particular MLST clades rather than with particular medical
4 conditions [17]. For example, one previous study identified an enrichment of MLST clade
5 1 *C. albicans* isolates recovered from sub-gingival sites in patients with untreated
6 periodontal disease; 21/31 isolates investigated belonged to MLST clade 1 and all were
7 found to belong to ABC genotype A [19]. This finding was not unusual, as many previous
8 studies have shown that MLST clade 1 consists predominantly of ABC genotype A
9 isolates [17]. No ABC genotype C isolates were identified in the periodontal study. In a
10 separate study on 14 oral *C. albicans*-positive patients with Autoimmune
11 Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy (APECED) syndrome, a rare
12 genetic disease characterised by autoimmunity to endocrine organs, ectodermal disorders
13 and chronic mucocutaneous candidiasis, 9/14 patients yielded genotype A isolates, six of
14 which belonged to MLST clade 1 [30]. One patient yielded genotype B isolates (MLST
15 clade 8) and four patients yielded genotype C isolates, two of which belonged to MLST
16 clade 4, one to clade 15, whereas the remaining isolate was a singleton.

17 Based mainly on case reports, the potential for malignant transformation of CL is
18 well recognised [31,32], and *Candida* has previously been implicated as having an
19 aetiological role in initiation of carcinoma in leukoplakia lesions associated with
20 acetylaldehyde production [33,34]. A recent study reported that *C. albicans* adhesion,
21 tissue invasion and damage of epithelial cells is influenced by a combination of fungal
22 morphology and activity and epithelial cell type and stage of differentiation, suggesting
23 that epithelial cells differ in their susceptibility to *C. albicans* [35]. This study also showed
24 that *C. albicans* can invade oral epithelial cells by induced endocytosis and by active
25 penetration.

1 In the present study two cases of malignant transformation occurred in diagnosed
2 CL lesions, one infected with an MLST clade 1, ABC genotype A isolate and the other
3 infected with an MLST clade 3, ABC genotype C isolate, during the follow up period of
4 two years (Table 1). Overall, the results of the present study do not show enrichment of
5 specific *C. albicans* clonal lineages among isolates recovered from CL lesions although
6 the apparent increased association of ABC genotype C isolates with CL lesions is
7 intriguing. Our findings suggest that *C. albicans* invades leukoplakia lesions giving rise to
8 *Candida* leukoplakia and that lesional factors are probably more significant than the *C.*
9 *albicans* lineage. Larger studies are required to further investigate the role of *Candida* in
10 CL.
11

1 **ACKNOWLEDGMENTS**

2

3 We acknowledge Imperial College London for hosting the *C. albicans* MLST website and
4 the Wellcome Trust for funding the database and website. We thank Dr. M.E. Bougnoux
5 at the Institut Pasteur, Paris, France for her advice and assistance, and for the continuous
6 curation and management of the *C. albicans* MLST database.

7

1 Table 1. Oral leukoplakia patient details and *Candida* species recovered

Patient	Sex	Age	Lesion site	Denture wearer	Smoker	Degree of dysplasia	<i>Candida</i> density	cell	<i>Candida</i> species isolated
CL							Rinse cfu/ml	Swab cfu	
01	M	58	PAL	Yes	Yes	Moderate	40	420	<i>C. albicans</i>
02	F	52	BM	Yes	Yes	No	70	35	<i>C. albicans</i>
03	M	74	TON	Yes	No	Severe	0	415	<i>C. albicans</i>
04	M	38	BM	No	Yes	Moderate	157	42	<i>C. albicans</i>
05	M	60	BM	Yes	Yes	Moderate	C	380	<i>C. albicans</i>
06	F	77	TON	No	No	Mild	1450	120	<i>C. albicans</i>
07	M	73	AR	Yes	Yes	No	0	7	<i>C. albicans</i>
08 ^a	F	58	GIN	No	No	Severe (SCC)	0	2 480	<i>C. glabrata</i> <i>C. albicans</i>
09	F	47	BM	No	No	Severe	0	1525	<i>C. albicans</i>
10	M	55	BM	Yes	Yes	Mild	20	130	<i>C. albicans</i>
11	F	73	BM	Yes	Yes	Moderate	60 120 10 25	84 4 7	<i>C. dubliniensis</i> <i>C. albicans</i> <i>C. parapsilosis</i> <i>C. guilliermondii</i>
12	F	44	BM	No	Yes	Moderate	2	15	<i>C. albicans</i>
13	M	29	BM	Yes	No	Severe	225 60 35	441 9 1	<i>C. albicans</i> <i>C. tropicalis</i> <i>S. cerevisiae</i>
14 ^a	M	63	BM	No	Yes	Severe (SCC)	SC	166	<i>C. albicans</i>
15	F	64	BM	No	No	Severe	C	SC	<i>C. albicans</i>
16	F	57	BM	No	Yes	Severe	450	400	<i>C. albicans</i>
17	M	41	BM	No	Yes	Moderate	160	10	<i>C. albicans</i>
18	M	69	BM	No	Yes	Moderate	27	0	<i>C. albicans</i>
19	F	76	PAL	Yes	No	Severe	SC	200	<i>C. albicans</i>
20	M	72	PAL	No	No	Moderate	290	0	<i>C. albicans</i>
21	F	63	TON	No	No	Severe	C	SC	<i>C. albicans</i>
22	M	77	BM	Yes	Ex	Moderate	0	0	None
23	M	45	BM	Yes	Yes	Moderate	10	2	<i>C. albicans</i>
24	F	60	BM	No	No	Moderate	240	125	<i>C. albicans</i>
25	M	47	BM	No	Yes	Severe	C 130	440 6	<i>C. albicans</i> <i>C. glabrata</i>
26	M	63	BM	No	Yes	Moderate	SC	SC	<i>C. albicans</i>
27	F	60	BM	Yes	Yes	Moderate	280	92	<i>C. albicans</i>
28	F	46	BM	Yes	Yes	Mild	0	250	<i>C. albicans</i>
29	M	46	TON	No	Yes	No	293	220	<i>C. albicans</i>
30	F	40	BM	No	Yes	Moderate	215	185	<i>C. albicans</i>
31	M	56	BM	Yes	Yes	Mild	0	0	None
NCL									
32	F	79	BM	Yes	Ex	Severe	20	C	<i>C. albicans</i>
33	M	55	BM	Yes	Ex	Severe	50	0	<i>C. albicans</i>
34	M	62	BM	No	Yes	Moderate	0	2	<i>C. albicans</i>
35	M	46	TON	No	Ex	Moderate	150	0	<i>C. albicans</i>
36	F	82	AR	Yes	Yes	Moderate	30 10	0 7	<i>C. albicans</i> <i>C. glabrata</i>
37	F	41	BM	No	Yes	Severe	48	18	<i>C. albicans</i>
38	M	83	TON	Yes	Yes	Moderate	172	150	<i>C. albicans</i>
39	M	68	PAL	No	Yes	No	4	0	<i>C. albicans</i>
40	F	39	BM	No	No	Moderate	10	0	<i>C. albicans</i>
41	M	59	TON	No	Ex	Severe	16	9	<i>C. albicans</i>

42	F	49	TON	No	No	Severe	SC	0	<i>C. albicans</i>
43	F	44	BM	No	No	Severe	83	32	<i>C. albicans</i>
44	M	51	PAL	No	Yes	Moderate	79	8	<i>C. albicans</i>
45	F	64	BM	No	Yes	Severe	39	10	<i>C. albicans</i>
46	M	76	BM	No	Yes	Mild	10	9	<i>C. albicans</i>
47	F	55	AR	No	Yes	Mild	50	11	<i>C. albicans</i>
48	M	40	AR	No	Yes	Mild	0	3	<i>C. albicans</i>
49	F	35	FM	Yes	Yes	Moderate	485	0	<i>C. albicans</i>
							113	1	<i>C. krusei</i>
50	F	65	BM	No	Yes	Moderate	20	0	<i>C. albicans</i>
51	M	67	BM	Yes	Yes	Moderate	0	0	None
52	F	62	FM	No	Yes	Moderate	100	0	<i>C. albicans</i>
53	M	73	GIN	No	Yes	Moderate	0	0	None
54	M	58	FM	No	Ex	Moderate	110	13	<i>C. albicans</i>
55	M	78	BM	No	Yes	Moderate	40	12	<i>C. albicans</i>
56	F	51	BM	No	Ex	Moderate	0	203	<i>C. albicans</i>
57	F	55	TON	Yes	No	Severe	260	107	<i>C. albicans</i>
58	F	87	AR	Yes	No	Mild	108	0	<i>C. albicans</i>
59	M	66	BM	No	Ex	No	3040	0	<i>C. albicans</i>
60	F	66	BM	No	Yes	Mild	20	0	<i>C. albicans</i>
61	M	60	PAL	No	Yes	Mild	477	16	<i>C. albicans</i>
62	F	78	BM	Yes	Yes	No	150	41	<i>C. albicans</i>
63	F	60	TON	Yes	Yes	Moderate	40	0	<i>C. albicans</i>
64	M	38	TON	No	Ex	Severe	120	60	<i>C. albicans</i>
65	M	47	BM	No	Yes	Moderate	180	5	<i>C. albicans</i>
66	F	48	PAL	No	Yes	Moderate	43	2	<i>C. albicans</i>
67	M	81	BM	Yes	Ex	Severe	C	270	<i>C. albicans</i>
							SC	120	<i>C. tropicalis</i>
68	F	73	BM	No	Yes	Severe	0	0	None
69	F	56	TON	No	Yes	Severe	0	0	None
70	F	62	BM	No	No	Severe	0	0	None
71	F	56	FM	No	No	Severe	0	0	None
72	M	62	AR	No	No	Severe	0	0	None
73	F	61	TON	No	No	Moderate	0	0	None
74	F	61	PAL	No	No	Severe	0	0	None
75	F	40	AR	No	No	No	0	0	None
76	M	51	GIN	No	No	No	0	0	None
77	M	51	BM	No	No	Severe	0	0	None
78	M	54	BM	No	No	Severe	0	0	None

1 ^aBoth of these patients with CL went on to develop squamous cell carcinoma at the lesional site within a
2 two-year follow up period.

3 Abbreviations; CL, *Candida* leuoplakia; NCL, Non-*Candida* leukoplakia; SCC, squamous cell carcinoma;
4 SC, semi-confluent growth (approx. 1000 cfu per swab or per ml oral wash); C, confluent (approx. 5000 cfu
5 per swab or per ml oral wash); BM, buccal mucosa; TON, tongue, PAL, palate, AR, alveolar ridge; FM,
6 floor of mouth; GIN, gingivae.

7

1 **Table 2. MLST DSTs, allelic profiles and ABC genotypes of *C. albicans* isolates**
 2 **recovered from OL patients and healthy oral carriers**

Isolate ^a	Source	DST	MLST loci							MLST clade ^b	CC	Htz	ABC Genotype ^c
			<i>AAT</i>	<i>ACC</i>	<i>ADP</i>	<i>MPI</i>	<i>SYA</i>	<i>VPS13</i>	<i>ZWF</i>				
CL10	CL	1235	2	2	2	31	2	24	5	1	1	14	A
CL09	CL	1234	2	2	2	27	2	24	5	1	1	14	A
CL29	CL	1659	2	2	2	2	2	24	5	1	1	15	A
CL01	CL	408	2	5	5	2	2	20	5	1	1	17	A
CL20	CL	1400	2	5	5	31	2	6	5	1	1	22	A
CL08 ^d	CL	1239	2	23	5	2	2	63	5	1	S	18	A
CL15	CL	1237	73	5	5	9	2	127	20	1	S	7	A
CL23	CL	228	8	5	20	6	43	69	22	8	23	14	A
CL25	CL	538	62	12	21	1	6	30	4	11	5	23	A
CL12	CL	2105	6	3	37	2	38	46	237	7	16	27	A
CL122	CL	2105	6	3	37	2	38	46	237	7	16	27	A
CL04	CL	149	2	5	5	2	2	21	20	1	1	18	A
CL17	CL	1240	13	10	15	24	50	37	15	3	4	21	B
CL21	CL	1294	21	26	14	18	72	86	84	6	8	21	B
CL22	CL	1682	4	31	6	4	61	15	201	15	58	11	B
CL16	CL	2106	13	10	15	4	7	32	15	3	4	21	B
CL24	CL	79	2	5	5	9	2	6	5	1	1	17	C
CL13	CL	1236	13	32	15	6	7	140	15	3	4	28	C
CL14 ^d	CL	1681	13	10	83	6	84	32	15	3	S	19	C
CL05	CL	1232	14	3	6	4	56	3	8	4	S	7	C
CL02	CL	1432	4	26	61	4	34	60	202	4	S	8	C
CL06	CL	1233	33	7	38	31	78	122	15	8	S	19	C
CL07	CL	1680	28	14	38	102	31	47	15	8	S	17	C
CL03	CL	1431	4	35	6	4	58	15	200	15	S	6	C
CL11	CL	2107	4	13	6	4	61	29	201	15	30	12	C
OL62	NCL	1648	8	3	2	4	2	6	5	1	1	9	A
OL65	NCL	1047	8	2	5	2	2	6	5	1	1	16	A
OL66	NCL	1651	2	2	5	2	2	63	20	1	1	8	A
OL32	NCL	69	2	5	5	2	2	6	5	1	1	23	A
OL64	NCL	1650	3	3	5	3	57	3	6	1	10	10	A
OL47	NCL	1238	35	2	4	4	49	4	4	2	2	3	A
OL61	NCL	1496	35	2	4	4	49	26	4	2	2	7	A
OL34	NCL	481	8	14	8	4	2	3	8	4	3	4	A
OL57	NCL	1683	62	3	3	3	26	16	95	9	6	19	A
OL59	NCL	1685	67	3	10	1	6	8	20	11	S	8	A
OL37	NCL	2108	35	4	4	4	4	41	4	2	3	19	A
OL38	NCL	37	2	5	5	2	2	21	5	1	1	22	A
OL55	NCL	1224	2	2	2	2	2	20	5	1	1	16	A
OL58	NCL	1684	28	14	38	2	106	122	15	8	S	22	B
OL39	NCL	1429	3	3	3	3	67	16	95	9	S	13	B
OL43	NCL	2109	4	82	6	4	61	32	201	15	S	11	B

OL63	NCL	1649	8	3	2	2	2	6	5	1	1	16	C
OL33	NCL	1686	11	8	4	3	7	19	4	5	S	23	C
OL35	NCL	1652	32	3	43	3	3	32	94	9	S	15	C
OC102	OC	1660	3	5	5	2	2	76	5	1	1	18	A
OC100	OC	1428	2	23	5	102	2	20	20	1	S	10	A
OC107	OC	1664	4	5	6	2	2	106	5	1	S	16	A
OC116	OC	1673	2	3	10	2	2	94	2	1	S	16	A
OC106	OC	1663	4	5	4	4	139	26	4	2	2	11	A
OC119	OC	275	35	2	4	4	4	4	4	2	2	8	A
OC103	OC	1661	40	24	41	21	4	76	27	2	S	13	A
OC109	OC	1666	4	2	14	4	139	41	67	2	S	7	A
OC112	OC	1669	36	2	6	4	49	41	4	2	S	4	A
OC110	OC	1667	8	14	8	4	56	10	8	4	3	9	A
OC115	OC	1672	4	60	6	4	4	41	4	2	S	11	A
OC117	OC	1674	13	3	6	34	62	8	47	5	S	16	A
OC108	OC	1665	5	27	37	4	34	105	12	11	S	20	A
OC114	OC	1671	4	19	6	4	61	15	201	15	58	8	A
OC301	OC	2110	35	4	4	4	49	4	4	2	3	11	A
OC303	OC	2111	2	3	6	2	2	5	5	1	1	19	A
OC304	OC	2108	35	4	4	4	4	41	4	2	3	19	A
OC305	OC	2112	2	5	5	2	2	5	12	1	1	17	A
OC306	OC	492	6	3	37	2	38	32	12	7	16	20	A
OC307	OC	1118	55	3	4	54	6	45	15	8	12	11	A
OC310	OC	2113	60	7	21	1	50	11	15	11	11	15	A
OC311	OC	1133	2	5	5	2	2	5	5	1	1	22	A
OC312	OC	24	2	5	5	2	2	24	5	1	1	18	A
OC313	OC	2114	35	4	4	4	4	4	26	2	3	15	A
OC315	OC	171	8	3	6	2	2	6	5	1	1	16	A
OC316	OC	444	2	5	5	4	2	6	5	1	1	17	A
OC105	OC	659	11	26	6	4	34	60	119	4	9	10	B
OC104	OC	1662	3	26	6	4	34	60	55	4	S	4	B
OC111	OC	1668	14	14	30	4	56	3	8	4	S	8	B
OC302	OC	2115	4	35	6	4	61	15	5	15	62	12	B
OC309	OC	299	4	17	21	19	27	83	22	12	7	15	B
OC113	OC	1670	8	7	6	4	56	3	118	4	S	5	C
OC308	OC	924	8	14	6	4	7	3	8	4	2	9	C
OC314	OC	2116	60	7	21	1	7	55	15	11	11	26	C

1 ^aWith regard to isolates recovered from CL and NCL patients, the numeral part of the isolate name refers to
2 the corresponding patient numbers shown in Table 1, column one (e.g. isolate CL10 was recovered from
3 patient 10). Isolates CL12 and CL122 were recovered from two separate CL lesions in patient 12. CL and
4 NCL isolates were recovered from lesional swabs, whereas OC isolates were recovered from oral rinse
5 samples.

6 ^bMLST clades defined according to Odds *et al.* [17].

7 ^cABC genotypes were assigned based on the presence or absence of an intron in the 25S rRNA gene [14].

1 ^dTwo CL patients included in this study (patients 08 and 14, Table 1) developed squamous cell carcinoma at
2 the lesional site within a two-year follow up period. Isolate CL14 (patient 14) and isolate CL08 (patient 8)
3 belonged to ABC genotypes C and A, respectively.

4 Abbreviations: DST, diploid sequence type; CC, clonal complex; Htz, number of heterozygous sites; CL,
5 *Candida* leukoplakia; NCL, non-*Candida* leukoplakia; OC, oral carriage; S, singleton.

6

1 **Table 3. Distribution of ABC genotypes in isolates recovered from oral leukoplakia**
 2 **patients and healthy carriers**

Patient Group	Number of isolates	Genotypes identified ^a n (%)
CL	25	A =12 (48) B =4 (16) C =9 (36)
NCL	19	A =13 (68.4) B =3 (15.8) C =3 (15.8)
OC	34	A=26 (76.5) B =5 (14.7) C =3 (8.8)

3 ^aThe prevalence of genotype A and genotype C isolates differed significantly between
 4 the CL and OC groups (P<0.02).

5 Abbreviations: CL, *Candida* leukoplakia; NCL, non-*Candida* leukoplakia; OC, oral
 6 carriage.

7
 8
 9

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

REFERENCES:

1. Odds FC. (1988) Ecology of *Candida* and epidemiology of candidosis. In: Odds FC editor. *Candida* and Candidosis. London: Baillière Tindall. pp. 68-92.
2. Warnakulasuriya S, Johnson NW, van der Waal I (2007). Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J Oral Pathol Med* 36: 575-580.
3. Lehner T (1964). Chronic Candidiasis. *Trans St Johns Hosp Dermatol Soc* 50: 8-21.
4. Cawson RA, Lehner T (1968). Chronic hyperplastic candidiasis--Candidal leukoplakia. *Br J Dermatol* 80: 9-16.
5. Krogh P, Holmstrup P, Thorn JJ, Vedtofte P, Pindborg JJ (1987). Yeast species and biotypes associated with oral leukoplakia and lichen planus. *Oral Surg Oral Med Oral Pathol* 63: 48-54.
6. Sittheequa MA, Samaranayake LP (2003). Chronic hyperplastic candidosis/candidiasis (Candidal leukoplakia). *Crit Rev Oral Biol Med* 14: 253-267.
7. Jepsen A, Winther JE (1965). Mycotic infection in oral leukoplakia. *Acta Odontol Scand* 23: 239-256.
8. Holmstrup P, Bessermann M (1983). Clinical, therapeutic, and pathogenic aspects of chronic oral multifocal candidiasis. *Oral Surg Oral Med Oral Pathol* 56: 388-395.
9. McCullough M, Jaber M, Barrett AW, Bain L, Speight PM, et al. (2002). Oral yeast carriage correlates with presence of oral epithelial dysplasia. *Oral Oncol* 38: 391-393.
10. Bartie KL, Williams DW, Wilson MJ, Potts AJ, Lewis MA (2001). PCR fingerprinting of *Candida albicans* associated with chronic hyperplastic candidosis and other oral conditions. *J Clin Microbiol* 39: 4066-4075.

- 1 11. Bougnoux ME, Morand S, d'Enfert C (2002). Usefulness of multilocus sequence
2 typing for characterization of clinical isolates of *Candida albicans*. J Clin Microbiol 40:
3 1290-1297.
- 4 12. Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, et al. (2003)
5 Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J
6 Clin Microbiol 41: 5265-5266.
- 7 13. Tavanti A, Gow NA, Senesi S, Maiden MC, Odds FC (2003). Optimization and
8 validation of multilocus sequence typing for *Candida albicans*. J Clin Microbiol 41:
9 3765-3776.
- 10 14. McCullough MJ, Clemons KV, Stevens DA (1999). Molecular and phenotypic
11 characterization of genotypic *Candida albicans* subgroups and comparison with *Candida*
12 *dubliniensis* and *Candida stellatoidea*. J Clin Microbiol 37: 417-421.
- 13 15. Odds FC, Davidson AD, Jacobsen MD, Tavanti A, Whyte JA, et al. (2006) *Candida*
14 *albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus
15 sequence typing. J Clin Microbiol 44: 3647-3658.
- 16 16. McCullough MJ, Clemons KV, Stevens DA (1999) Molecular epidemiology of the
17 global and temporal diversity of *Candida albicans*. Clin Infect Dis 29: 1220-1225.
- 18 17. Odds FC, Bougnoux ME, Shaw DJ, Bain JM, Davidson AD, et al. (2007) Molecular
19 phylogenetics of *Candida albicans*. Eukaryot Cell 6: 1041-1052.
- 20 18. Walker DM, Arendorf TM (1990) Candidal leucoplakia, chronic multifocal candidosis
21 and median rhomboid glossitis. In: Samaranayake LP, MacFarlane TW editors. Oral
22 Candidosis. Guilford, UK: John Wright, pp. 184-199.
- 23 19. McManus BA, Maguire R, Cashin PJ, Claffey N, Flint S, et al. (2012) Enrichment of
24 multilocus sequence typing clade 1 with oral *Candida albicans* isolates in patients with
25 untreated periodontitis. J Clin Microbiol 50: 3335-3344.

- 1 20. Odds FC, Bernaerts R (1994) CHROMagar Candida, a new differential isolation
2 medium for presumptive identification of clinically important *Candida* species. J Clin
3 Microbiol 32: 1923-1929.
- 4 21. Coleman DC, Sullivan DJ, Bennett DE, Moran GP, Barry HJ, et al. (1997)
5 Candidiasis: the emergence of a novel species, *Candida dubliniensis*. AIDS 11: 557-567.
- 6 22. Pincus DH, Coleman DC, Pruitt WR, Padhye AA, Salkin IF, et al. (1999) Rapid
7 identification of *Candida dubliniensis* with commercial yeast identification systems. J
8 Clin Microbiol 37: 3533-3539.
- 9 23. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular
10 Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150-163.
- 11 24. Odds FC, Jacobsen MD (2008). Multilocus sequence typing of pathogenic *Candida*
12 species. Eukaryot Cell 7: 1075-1084.
- 13 25. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring
14 patterns of evolutionary descent among clusters of related bacterial genotypes from
15 multilocus sequence typing data. J Bacteriol 186: 1518-1530.
- 16 26. Spratt BG, Hanage WP, Li B, Aanensen DM, Feil EJ (2004). Displaying the
17 relatedness among isolates of bacterial species -- the eBURST approach. FEMS Microbiol
18 Lett 241: 129-134.
- 19 27. Bougnoux ME, Diogo D, François N, Sendid B, Veirmeire S, et al. (2006) Multilocus
20 sequence typing reveals intrafamilial transmission and microevolutions of *Candida*
21 *albicans* isolates from the human digestive tract. J Clin Microbiol 44: 1810-1820.
- 22 28. Chen KW, Chen YC, Lo HJ, Odds FC, Wang TH, et al. (2006) Multilocus sequence
23 typing for analyses of clonality of *Candida albicans* strains in Taiwan. J Clin Microbiol
24 44: 2172-2178.

- 1 29. Wrobel L, Whittington JK, Pujol C, Oh SH, Ruiz MO, et al. (2008) Molecular
2 phylogenetic analysis of a geographically and temporally matched set of *Candida*
3 *albicans* isolates from humans and nonmigratory wildlife in central Illinois. *Eukaryot Cell*
4 7: 1475-1486.
- 5 30. McManus, B.A., McGovern, E., Moran, G.P., Healy, C.M., Nunn, J., et al. (2011)
6 Microbiological screening of Irish patients with autoimmune polyendocrinopathy-
7 candidiasis-ectodermal dystrophy reveals persistence of *Candida albicans* strains, gradual
8 reduction in susceptibility to azoles, and incidences of clinical signs of oral candidiasis
9 without culture evidence. *J Clin Microbiol* 49: 1879-1889.
- 10 31. Cawson RA (1966) *Chronic* oral candidiasis and leukoplakia. *Oral Surg Oral Med*
11 *Oral Pathol* 22: 582-591.
- 12 32. Williamson DM (1969) Chronic hyperplastic candidiasis and squamous carcinoma. *Br*
13 *J Dermatol* 81: 125-127.
- 14 33. Krogh P (1990) The role of yeasts in oral cancer by means of endogenous nitrosation.
15 *Acta Odontol Scand* 48: 85-88.
- 16 34. Gainza-Cirauqui ML, Nieminen MT, Novak Frazer L, Aguirre-Urizar JM, Moragues
17 MD, et al. (2013) Production of carcinogenic acetaldehyde by *Candida albicans* from
18 patients with potentially malignant oral mucosal disorders. *J Oral Pathol Med* 42: 243-
19 249.
- 20 35. Dalle F, Wächtler B, L'Ollivier C, Holland G, Bannert N, Wilson D, Labruère C,
21 Bonnin A, Hube B (2010) Cellular interactions of *Candida albicans* with human oral
22 epithelial cells and enterocytes. *Cell Microbiol* 12:248-271.
23

1 **Figure 1. UPGMA dendrogram depicting the genetic relatedness of all *C. albicans***
2 **isolates subjected to MLST and ABC genotyping analysis in the present study.**

3 Individual isolates recovered from CL (n=25) and NCL (n=19) patients are indicated with
4 the letters CL and OL in the isolate names, respectively. The numeral part of the isolate
5 name refers to the corresponding patient numbers as shown in Table 1, column 1. Isolates
6 recovered from age and sex-matched healthy control patients (n=34) by oral rinse are
7 indicated using the letters OC in the isolate name. Hyphenated letters A, B or C following
8 each isolate name indicate ABC genotypes. The scale bar indicates p- distance. Distinct
9 MLST clades are indicated by separated square parenthesis to the right of the isolate
10 names and labeled with previously designated clade numbers [17]. Numbers at clade
11 branches indicate bootstrap support levels, based on 1000 replications. Overall, the
12 population analysis based on MLST suggests no clonal enrichment of *C. albicans* isolates
13 recovered from CL or NCL lesions. Isolates recovered from CL patients are distributed
14 among eight clades, isolates from NCL patients are distributed among eight clades and
15 isolates from healthy carriers are distributed among nine clades, with clade 1
16 predominating in all three groups.

17

