

A Dysbiotic Mycobiome Dominated by *Candida albicans* is Identified within Oral Squamous Cell Carcinomas

Manosha Perera ^{1,†}, Nezar Noor Al-hebshi ^{2,†,*}, Irosha Perera ³, Deepak Ipe ^{4,5}, Glen Ulett ^{4,5}, David J. Speicher ^{5,6,7}, Tsute Chen ⁸ and Newell W. Johnson ^{1,5}

¹ School of Dentistry and Oral Health, Griffith University, Nathan QLD 4111, Australia; manosha.perera@griffithuni.edu.au (M.P.), n.johnson@griffith.edu.au (N.W.J.)

² Maurice H. Kornberg School of Dentistry, Temple University, Philadelphia, PA 19122, USA

³ Preventive Oral Health Unit, The National Dental Hospital (Teaching), Colombo, Sri Lanka; irosha_rukmali@yahoo.com

⁴ School of Medical Science, Griffith University, Nathan QLD 4111, Australia; d.ipe@griffith.edu.au (D.I.); g.ulett@griffith.edu.au (G.U.)

⁵ Menzies Health Institute Queensland, Griffith University, Nathan QLD 4111, Australia; d.speicher@griffith.edu.au

⁶ Microbiology and Infectious Diseases, St. Joseph's Healthcare Hamilton, Hamilton, ON L8N 4A6, Canada

⁷ Department of Pathology & Molecular Medicine, McMaster University, Hamilton, ON L8S 4L8, Canada

⁸ Department of Microbiology, Forsyth Institute, Cambridge, MA 02142, USA

† Both authors contributed equally to the manuscript

*Corresponding author

Nezar Noor Al-hebshi, Department. Mobile: +966-565288886, e-mail address: Maurice H.

Kornberg School of Dentistry, Temple University, Philadelphia, USA. e-mail address:

alhebshi@temple.edu

Running title: *Mycobiome associated with OSCC*

Abstract

Background: Studies employing next-generation sequencing (NGS) show that the oral fungal community (mycobiome) is far more complex than hitherto thought. However, the role of the oral mycobiome in health and disease, including oral carcinogenesis, has not been explored. **Objective:** To characterize the mycobiome associated with oral squamous cell carcinoma (OSCC). **Methods:** Tissue biopsies [cases: 25 OSCC; controls: 27 intra-oral fibro-epithelial polyp (FEP)] were collected from oral and maxillofacial units in Sri Lanka. Total DNA was extracted and subjected to sequencing of the fungal ITS2 region using Illumina's 2x300 bp chemistry. High quality, non-chimeric merged reads were classified to species level using a BLASTN-algorithm with UNITE's named species sequences as reference. Downstream analyses were performed using QIIME and LefSe. **Results:** 364 species representing 160 genera and 2 phyla (Ascomycota and Basidiomycota) were identified, with *Candida* and *Malassezia* making up 48% and 11% of the average mycobiome, respectively. However, only 5 species and 4 genera were detected in $\geq 50\%$ of the samples. The species richness and diversity were significantly lower in OSCC. At the genus level, *Candida*, *Hannaella* and *Gibberella* were overrepresented in OSCC while *Alternaria* and *Trametes* were more abundant in FEP. Species-wise, *C. albicans*, *C. etchellsii* and *Hannaella luteola*-like species were enriched in OSCC while *Malassezia restricta*, *Aspergillus tamaraii*, *Alternaria alternate*, *Cladosporium halotolerans*, and *Hanseniaspora uvarum*-like species were the most significantly abundant in FEP. **Conclusions:** A dysbiotic mycobiome dominated by *C. albicans* was found in association with OSCC. Whether this dysbiosis plays a role in oral carcinogenesis warrants further investigation.

Key words: 18S rRNA; fungi; mycobiome; carcinoma; High-Throughput Nucleotide Sequencing; microbiome; mouth; squamous cell

Introduction

Oral squamous cell carcinoma (OSCC) accounts for more than 90% of cases of oral cancer, a malignancy with poor prognosis representing a public health challenge, particularly in less developed regions where it ranks the 8th most common cancer type [1, 2, 3]. The incidence rates are exceptionally high in certain countries such as Papua New Guinea, India, Sri Lanka, Maldives and Pakistan [1]. There are a number of well-established risk factors for OSCC including smoking, smokeless tobacco usage, areca nut and betel quid chewing, alcohol consumption and HPV infections [4, 5]. However, around 15% of oral cancer cases are not attributable to any of these; the role of other factors such as infection and inflammation is thus emerging [6, 7].

The oral cavity harbors the second most diverse microbiota in our body: the oral microbiome [8]. Oral microorganisms, predominantly bacteria, form homeostatic communities that live in a mutualistic relationship with the host. However, local ecological challenges may disturb the community balance and result in a microbial “dysbiosis” characterized by altered microbiome profile and the potential to cause or contribute to disease [9]. Technological advances, particularly the advent of next-generation sequencing (NGS), have revolutionized the study of oral microbiome in health and disease [10]. Periodontitis and dental caries, the most common oral diseases, are recognized today as sequelae of oral microbial dysbiosis [9, 11]. Recent evidence from NGS studies also suggest a possible role of a dysbiotic microbial community in OSCC [7, 12]. However, all these studies have focused only on the bacterial component of the microbiome (i.e. bacteriome).

Numerous fungi are found in the oral cavity, with the genus *Candida* being the most commonly isolated and thus studied. Recent studies using NGS have revealed the presence of a complex fungal community (mycobiome) in healthy individuals. In 2010, a study by Ghannoum et al. revealed

presence of a core oral mycobiome consisting of 13 taxa, with *Candida*, Saccharomycetales, *Cladosporium* and *Aspergillus* being the most abundant [13]. In 2014, Dupuy et al. [14] identified an additional 5 core genera, of which *Malassezia* was detected for the first time in the oral cavity and even found to be more abundant than *Candida* in some of the subjects. So far, no attempts have been made to explore the potential role of the oral mycobiome in oral health and disease, including oral cancer. This is surprising given the existing evidence implicating *Candida albicans* in oral carcinogenesis [15].

The objective of this study was to use NGS coupled with a species-level taxonomy assignment algorithm to compare the mycobiome profile within OSSC tissues to benign intra-oral fibro-epithelial polyps (FEP) and identify fungal taxa that may play a role in oral carcinogenesis.

Methods

Study design, setting and subjects

This was a case-control study. Study subjects were recruited between 17/04/15 and 02/08/15 at nine oral and maxillofacial (OMF) units in six provinces of Sri Lanka. Cases comprised 25 Sinhala, \geq 40 years old males with histologically confirmed OSCC affecting the buccal mucosa or tongue. The control group consisted of 27 Sinhala males with a clinical diagnosis of FEP also involving the buccal mucosa or tongue. Subjects with history of antibiotic used in the last 2 months were excluded.

Ethical approval of the study was obtained from the Faculty Research committee, Faculty of Dental Sciences, University of Peradeniya, Sri Lanka (FRC/FDS/UOP/E/2014/32) and Griffith University

Human Research Ethics Committee, Australia (DOH/18/14/HREC). Written informed consent was obtained from each participant.

History taking and clinical examination

Data were collected using a pre-tested, interviewer-administered questionnaire which comprised information on socio-demographics and risk habits including betel quid chewing, smoking and alcohol consumption. Clinical oral examinations were conducted by dental public health specialists. The oral mucosa was thoroughly inspected for any growth, ulceration or white patches. Number of missing teeth was recorded. Oral hygiene status was assessed with the simplified oral hygiene index [16] while periodontal status was assessed using bleeding on probing (BOP), periodontal pocket depth (PPD) and clinical attachment loss (CAL) at 4 sites per anterior tooth and 6 sites per posterior tooth.

Tissue sampling and DNA extraction

For suspected OSSC cases, tissue samples were obtained from the incisional biopsies taken for diagnosis. The freshly-taken biopsy was laid on a pile of sterile gauze and a small piece of tissue ($\sim 3 \text{ mm}^3$) was excised from the body of the lesion. A new sterile surgical blade was used for each case. The sample was aseptically transferred into a screw-cap vial and immediately stored at -20°C . Simultaneously, the rest of the biopsy was sent in 10% buffered formalin for histopathological diagnosis. Only samples histopathologically-confirmed as OSSC were included in the study. Control tissue samples were obtained as above from freshly excised, clinically diagnosed FEPs.

Tissue samples (~100 mg each) were finely chopped using a sterile blade. DNA extraction was then performed using Gentra Puregene Tissue kit (Qiagen, Germany) according to the manufacturer's instructions (solid tissue protocol) with a few modifications: 1) incubation in the lysis buffer was performed overnight; 2) an additional lysis step using 50 units of mutanolysin at 37°C for 1.5 hrs to digest the cell wall of Gram +ve bacteria was included prior to the addition of Gentra Puregene Proteinase K. Total DNA concentration and purity was determined using the NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific, USA). The extracts were stored at - 80°C.

Amplicon Library preparation and sequencing

Amplicon library was prepared as per Illumina's protocol except that amplification was performed with the Q5 Hot Start High-Fidelity 2X MasterMix (New England Biolabs, USA). In brief, the fungal ribosomal internal transcribed spacer 2 (ITS2) was amplified using the primers ITS3-F (5'-GCATCGATGAAGAACGCAGC-3') and ITS4-R (5'-TCCTCCGCTTATTRATATGC-3') [17], linked to Illumina's specific adapter sequences 803F and 1392wR, in standard PCR conditions. The resultant PCR amplicons (~ 250-590 bp) were purified using Agencourt AMPure XP beads (Beckman Coulter, USA). A second PCR performed to tag the amplicons with unique 8-base barcodes using Nextera XT v2 Index Kit sets A-D (Illumina, USA). The tagged amplicons were then pooled together in equimolar concentrations and sequenced on MiSeq Sequencing System (Illumina, USA) using v3 2x300 bp, paired-end sequencing chemistry in the Australian Centre for Ecogenomics according to the manufacturer's protocol.

Processing of sequencing data

Raw sequencing data were deposited in (and are publicly available from) Sequence Reads Archive (SRA) under project no. PRJNA375780. Reads with primer mismatches were removed and primer sequences were trimmed off. Paired sequences were then merged with PEAR [18] using the following parameters: minimum amplicon length, 213 bp; maximum amplicon length, 552 bp; p-value, 0.001. Preprocessing of the merged reads was performed using the mothur v1.38.1 [19]. Firstly, to stringently minimize sequencing errors, reads with ambiguous bases, with homopolymers > 8 bp or that did not achieve a sliding 50-nucleotide Q-score average of ≥ 30 were filtered out. Secondly, the high quality reads were cleared of chimeras with Uchime [20] using the self-reference approach [21]. Finally, sequences representing non-fungal lineages, identified by preliminary taxonomy using mothur's classify.seqs command, were removed.

Taxonomy assignment algorithm and down-stream analysis

The high quality, non-chimeric reads were classified to the species level employing a previously described BLASTN-based algorithm, modified to analyze fungal ITS2 region of the 18S rRNA gene instead of bacterial 16S RNA [22]. A set of 23,423 fungal ITS sequences representing all named species (16,595 species) in UNITE's database version 7.1 (<https://unite.ut.ee/repository.php>; August 22, 2016 dynamic release; untrimmed sequences) [23] was used as reference; the fasta and taxonomy files of this set can be downloaded at ftp://www.homd.org/publication_data/20170221/. Briefly, the reads were individually BLASTN-searched against the reference set at alignment coverage of $\geq 99\%$ and percent identity of $\geq 98.5\%$. Hits were ranked by percent identity and, when equal, by bit score. Reads were assigned taxonomies of the best hits. Reads with best hits representing more than one species were screened again for chimeras using *de novo* check at 98%

similarity with USEARCH v8.1.1861, and, if not chimeric, were assigned multiple-species taxonomy[24]. Reads with no matches at the specified criteria underwent secondary *de novo* chimera checking as above, and then *de novo*, species-level operational taxonomy unit (OTU) calling at 98% using USEARCH. Singleton OTUs were excluded; the rest were considered potentially novel species and a representative read from each was BLASTN-searched against the same reference sequence set again to determine the closest species for taxonomy assignment.

Downstream analysis was performed as previously described [22]. In short, the Quantitative Insights Into Microbial Ecology (QIIME) v1.9.1 [25] was employed to perform further analysis including generation of taxonomy plots, rarefaction, calculation of species richness and diversity indices, computing distance matrices, and running principle component analysis (PCoA). Detection of differentially abundant taxa between the cases and controls was done using Linear discriminant analysis Effect Size (LEfSe) [26].

Results

Four samples ended up with low read count (< 3000) and one with very high count (an outlier) and were thus excluded. Results are presented hereafter for 22 cases and 25 controls. The clinical characteristics of both groups is presented in **Table 1**. The cases, albeit older on average than controls, were heavier users of betel quid and alcohol, and smoked more: these being the traditional risk factors for oral cancer. They had significantly poorer oral hygiene, fewer teeth and more severe periodontal disease.

DNA extracts and sequencing/data processing statistics

Total DNA extracted averaged 482.08 ng/μL (range: 16.85-1908 ng/μL; <100 ng/μL from 8/47 (17%) samples) with all samples containing DNA of high purity (average 260/280 ratio was 1.93, range: 1.81-2.07). However, the mean 260/230 ratio was 1.54 (range: 0.41-2.22; 260) indicating presence of inhibitors in some of the samples, an issue that was dealt with using dilution before amplification.

Sequencing generated 1,576,427 raw paired reads: 14.3% of these were discarded due to primer mismatches; 96.6% of the remaining reads were successfully stitched with PEAR. Quality filtration and chimera checking removed 7.9% of the merged reads, thus leaving 1,063,430 reads (67.5%), 205-535 bp long. Of these, 1,017,131 reads (95.7%) were successfully classified to the species level; 2.3% did not return BLASTN matches and 2% formed singleton OTUs and were excluded. The number of classified reads per sample averaged 21,641 (range: 3,973 – 54,849).

Overall mycobiome profile

A total of 364 species belonging to 162 genera and 2 phyla were detected in the samples. The abundances and detection frequencies of these taxa in each of the samples and across the study groups are presented in **Supplementary Tables 1-3**. The number of species per sample ranged from 4 to 64. However, only 125 species and 74 genera were identified in more than one sample; 10 species and 7 genera in $\geq 25\%$ of the samples and 5 species and 4 genera in $>50\%$.

The abundances of the two phyla identified as well as the genera and species detected in $\geq 15\%$ of the samples are shown in **Figure 1**. On average, the phyla Ascomycota and Basidiomycota

accounted for 78.4% and 21.6% of the mycobiome, respectively. At the genus level, *Candida* was detected in 100% of the samples and constituted 48% of the average mycobiome. The genera *Malassezia*, *Cladosporium* and *Aspergillus* were identified in $\geq 75\%$ of the sample with an average abundance of 11%, 6.1% and 3.7%, respectively. Species-wise, *C. albicans* was found in all samples at a mean abundance of 44.4%. *Malassezia restricta*, *Aspergillus penicillioides* and *Malassezia globosa* were identified in 83%, 70.2% and 68.1% of the samples respectively, and accounted for 3.2%, 2.2% and 4.2% of the average mycobiome, respectively. *Cladosporium exasperatum* and a potentially novel species close to *Cladosporium sphaerospermum* were also identified in half of the samples at an average abundance of $> 2\%$.

However, the relative abundance of these taxa varied significantly between the samples. In addition, the mycobiome of some of the samples was dominated by taxa other than the above-mentioned. Examples of species abundant in single samples include *Rhodotorula mucilaginosa*, *Sporidiobolus johnsonii*, *Penicillium toxicarium*, *Toxicocladosporium irritans*, *Gibberella intricans*, *Alternaria infectoria*, *Ophiocordyceps sinensis* and *Aspergillus tamarii* as well as a number of potentially novel taxa.

OSCC vs. FEP

The number of species per sample ranged from 4 to 29 for the cases and from 8 to 64 for the controls. The FEP controls had higher species richness and α -diversity than the cases (**Table 2**). Rarefaction curves show that as few as 1,500 reads per sample represented sufficient sequencing depth (**Figure 2 A**). No separate clusters formed for the cases and controls by PCoA (**Figure 2 B**).

The genera and species found by LEsFe to be differentially abundant between the cases and controls are shown in **Figure 3**. The genera *Candida*, *Hannaella* and *Gibberella* were significantly more abundant in OSCC. In contrast, *Trametes* and *Alternaria* were strongly associated with FEP. At the species level, *C. albicans*, *Candida etchellii* and a potentially novel species close to *Hannaella luteola* were significantly enriched in OSCC. *C. albicans* was identified in 100% of the samples but the average relative abundance in OSCC was twice that in the controls (61.2% vs. 29.6%). *C. etchellsii* was identified in 32% of the cases compared to only 8% of the controls. On the other hand, a potentially novel *Hanseniaspora uvarum*-like species, in addition to *M. restricta*, *A. tamarii*, *Cladosporium halotolerans*, *Alternaria alternata* and *Malassezia furfur* were overrepresented or even exclusively found in FEP (See **Supplementary Table 4** for a list of taxa exclusively found in either group at prevalence $\geq 10\%$)

Discussion

To our knowledge, this his is the first study to explore the mycobiome associated with oral cancer. The study subjects were recruited from Sri Lanka, a developing lower-middle income country where oral cancer is the most common malignancy among men [1]. Tissue biopsies rather than surface swabs were used to ensure that any association identified is more relevant to carcinogenesis. Fibro-epithelial polyps (FEP) were chosen as controls with the assumption they would offer a more valid comparison (benign vs. malignant growths) than that with buccal epithelium or even adjacent normal tissue. However, some study limitations should be noted. Firstly, contamination of samples by saliva cannot be entirely excluded although great care was taken during biopsy to avoid this, and all biopsies were blotted with fresh gauze immediately after

excision. Secondly, some DNA samples had low 260/230 ratios, suggestive of the presence of inhibitors. Nevertheless, these samples amplified successfully after dilution. Finally, a few reads were detected in the control samples which, while having negligible effect on the results, suggests some sort of contamination at some stage of the laboratory work. The latter could have been due to contamination of the DNA extraction or sequencing reagents by extraneous organisms, now known to be a problem around the world [27].

Coupling Illumina's 2x300 bp sequencing chemistry and merging of paired reads ensured full ITS2 amplicons were obtained and maximizing taxonomic resolution. This, in addition to limiting the reference set to sequences from named species only, enabled us to classify the reads to the species level. A percent identity threshold of 98.5% (i.e. like that used to generate species hypothesis in UNITE's dynamic release) and high query alignment coverage ($\geq 98\%$) cutoff increased the reliability of taxonomic assignment and eliminated the possibility of forced classifications by BLASTN. However, this approach carries the risk of assigning reads belonging to species not represented in the reference set to another closely related species.

Of the species identified 239/364 (65.6%) were from single samples indicating that they represent transient environmental fungi. Although the scope of the current study is different from that of the previous two oral mycobiome studies in which salivary samples from healthy subjects were analyzed, some comparisons can be made. Ghannoum et al., using a relative abundance cutoff of 1%, identified 101 species in their samples [13]. Applying the same cutoff to our samples left 137 species, which is comparable. They also described a basal mycobiome comprising 13 taxa that was present in $\geq 20\%$ of the subjects. Consistently, 7 of these genera, namely *Candida*,

Aspergillus, *Cladosporium*, *Alternaria*, *Cryptococcus*, *Gibberella* and *Saccharomyces* were detected at the same frequency in at least one of our two groups. Interestingly, *Malassezia*, which was not detected by Ghannoum et al., was identified in this study as the second most common and abundant genus after *Candida*. This substantiates the more recent findings by Dupuy et al., who, for the first time, described this genus in the oral cavity and even found it to be more abundant than *Candida* [14]. The same study described a core mycobiome comprising 14 core genera, detected in $\geq 50\%$ of the subjects. Four of these (*Candida*, *Malassezia*, *Aspergillus*, *Cladosporium*) were also found in more than half of our samples. Put together, these findings provide evidence for the existence of a resident, core oral fungal community.

Candida, specifically *C. albicans* and *etchellsii*, showed the strongest association with OSCC in the current study. In fact, candidiasis has for long been proposed as a risk factor for malignant transformation of oral potentially malignant disorders, including leukoplakia [28]. Some strains of *C. albicans* have high nitrosation potential and have been experimentally shown to induce dysplasia [29]. Recently, there is increasing evidence of the association between candida infection and OSCC [15]. Nagy et al. recovered *C. albicans* from the surface of 8/21 (38%) neoplasms but from none of the control tissues [30]. Similarly, Čanković et al. isolated *Candida* from 9/30 (4 *C. albicans* and 5 non-*albicans*) cancer surfaces (30%) but from only 2/30 (6.6%) benign control samples [31]. Gall et al. identified *Candida spp* in 31/48 (65%) oral cancer cases, but the study did not include healthy tissue samples for comparison [32]. More recently, Berkovits et al. isolated yeasts, predominantly *Candida spp.*, from 18/20 (90%) OSCC samples compared to only 12/40 (30%) control tissues; in addition, they found the “fungal burden” to be higher in OSCC [33]. In our study, *Candida albicans* was detected in 100% of the samples, which is probably a reflection of the much higher sensitivity

of NGS compared to cultural techniques used in previous studies. The difference between the cases and controls in our material was rather in terms of relative abundance (average of 61.2% in the cases vs. 29.6% in the controls). This is somewhat consistent with the high “fungal burden” described by Berkovits et al. and indicates that a fungal community dysbiosis characterized, by overall simplification, with increased abundance of *C. albicans* is associated with OSCC.

C. etchellsii was identified in 32% of the cases vs. 8% of the controls: this is the first time a specific *Candida* species other than *C. albicans* is implicated in oral cancer, a finding worth further investigation. *Hannaella* and *Gibberella* spp were also overrepresented in OSCC tissues. These are typically found on plants as commensals and pathogens, respectively [34, 35]. Therefore, they probably represent contaminants of the OSCC rather than members of the oral mycobiome community, especially that they were found in only one control sample. However, their contribution to the carcinogenic process cannot be excluded; in fact one *Gibberella* species (*G. moniliformis*) produces a mycotoxin that has been shown to induce liver cancer in rodents [36].

Species belonging to *Malassezia*, *Aspergillus*, *Alternaria*, *Cladosporium* and *Hanseniaspora* were significantly enriched or exclusively found in the controls. *Malassezia* spp are normal colonizers of healthy skin [37] and have been recently found to be dominant members of the salivary mycobiome [14]. *Aspergillus*, *Alternaria*, and *Cladosporium* have also been described as core oral fungal taxa [13, 14]. Interestingly, some species of these genera, including *A. tamarii* and *A. alternata*, which were identified in our samples, are known to produce compounds with anticancer activity [38, 39, 40, 41]. In addition, statins produced by *A. tamarii* have been shown to inhibit growth of *C. albicans* [40], which possibly explains why all samples with high abundance of this

species had very low levels of *C. albicans* (**Supplementary Table 3**). Therefore, while these species may simply represent transient environmental fungi or passenger oral fungal taxa, it is also possible that carriage of some of these species confers some protection against development of oral cancer. Further research to explore these scenarios is needed to harness their potential for novel prevention and control strategies.

In conclusion, the current study uncovered a dysbiotic mycobiome characterized by lower species diversity and increased *C. albicans* abundance in association with OSCC. It also identified a number of fungal taxa capable of producing anti-cancer compounds exclusively in non-cancerous tissues. The relevance of these findings for development of risk markers and for prevention and control of OSCC warrants further investigation.

Acknowledgments

We acknowledge Dr WM Tilakaratne, Senior Professor of Oral Pathology, the Dean, Faculty of Dental Sciences, University of Peradeniya, Sri Lanka and Prof. L. Samaranayake, Honorary Professor of Oral Microbiomics & Infection, School of Dentistry, University of Queensland for valuable guidance. We thank Oral & Maxillo-Facial Surgeons- Dr Sharika Gunathilake, Dr S.A.K.J Kumara, Dr Ranjith Lal Kandewatte, Dr P. Kirupakaran, Dr D.K.Dias, Dr Chamara Athukorale, Dr Suresh Shanmuganathan and Dr T. Sabesan for facilitating data and sample collection from their respective Oral & Maxillo-Facial Units.

Competing interests

None to declare.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer.; 2013 [cited 2013]. Available from: <http://globocan.iarc.fr>.
2. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol.* 2009;45:309-16. Epub 2008/09/23.
3. Wang B, Zhang S, Yue K, Wang XD. The recurrence and survival of oral squamous cell carcinoma: a report of 275 cases. *Chin J Cancer.* 2013;32:614-8.
4. Petti S. Lifestyle risk factors for oral cancer. *Oral oncology.* 2009;45:340-50.
5. Gupta B, Johnson NW, Kumar N. Global Epidemiology of Head and Neck Cancers: A Continuing Challenge. *Oncology.* 2016;91:13-23.
6. Chocolatewala N, Chaturvedi P, Desale R. The role of bacteria in oral cancer. *Indian J Med Paediatr Oncol.* 2011;31:126-31. Epub 2011/05/18.
7. Perera M, Al-Hebshi NN, Speicher DJ, Perera I, Johnson NW. Emerging role of bacteria in oral carcinogenesis: a review with special reference to perio-pathogenic bacteria. *J Oral Microbiol.* 2016;8:32762.
8. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486:207-14.
9. Kilian M, Chapple IL, Hannig M, Marsh PD, Meuric V, Pedersen AM, Tonetti MS, Wade WG, Zaura E. The oral microbiome - an update for oral healthcare professionals. *Br Dent J.* 2016;221:657-66.
10. Chen H, Jiang W. Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front Microbiol.* 2014;5:508. Epub 2014/10/30.
11. Wade WG. The oral microbiome in health and disease. *Pharmacol Res.* 2013;69:137-43.
12. Al-Hebshi NN. Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma (in press). *Scientific Reports.* 2017.
13. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS pathog.* 2010;6:e1000713.
14. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, Dongari-Bagtzoglou A, Diaz PI, Strausbaugh LD. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of *Malassezia* as a prominent commensal. *PLoS One.* 2014;9:e90899.
15. Mohd Bakri M, Mohd Hussaini H, Rachel Holmes A, David Cannon R, Mary Rich A. Revisiting the association between candidal infection and carcinoma, particularly oral squamous cell carcinoma. *J Oral Microbiol.* 2010;2.
16. Greene JC, Vermillion JR. The Simplified Oral Hygiene Index. *J Am Dent Assoc.* 1964;68:7-13.
17. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol.* 2009;5:e1000352.
18. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics.* 2014;30:614-20. Epub 2013/10/22.
19. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75:7537-41. Epub 2009/10/06.
20. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011;27:2194-200. Epub 2011/06/28.

21. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*. 2011;6:e27310. Epub 2011/12/24.
22. Al-hebshi NN. Composition and imputed functions of the bacteriome in smokeless tobacco products with different oral carcinogenicity. *Genes*. 2017:In press.
23. Koljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Duenas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lucking R, Martin MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Poldmaa K, Saag L, Saar I, Schussler A, Scott JA, Senes C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol*. 2013;22:5271-7.
24. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460-1. Epub 2010/08/17.
25. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*. 2010;7:335-6. Epub 2010/04/13.
26. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.
27. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog*. 2016;8:24.
28. Sankari SL, Gayathri K, Balachander N, Malathi L. Candida in potentially malignant oral disorders. *Journal of pharmacy & bioallied sciences*. 2015;7:S162.
29. O'Grady JF, Reade PC. Candida albicans as a promoter of oral mucosal neoplasia. *Carcinogenesis*. 1992;13:783-6.
30. Nagy K, Sonkodi I, Szöke I, Nagy E, Newman H. The microflora associated with human oral carcinomas. *Oral oncology*. 1998;34:304-8.
31. Čanković M, Bokor-Bratić M. Candida albicans infection in patients with oral squamous cell carcinoma. *Vojnosanitetski pregled*. 2010;67:766-70.
32. Gall F, Colella G, Di Onofrio V, Rossiello R, Angelillo IF, Liguori G. Candida spp. in oral cancer and oral precancerous lesions. *New Microbiol*. 2013;36:283-8.
33. Berkovits C, Tóth A, Szenzenstein J, Deák T, Urbán E, Gácsér A, Nagy K. Analysis of oral yeast microflora in patients with oral squamous cell carcinoma. *SpringerPlus*. 2016;5:1257.
34. Landell MF, Brandao LR, Barbosa AC, Ramos JP, Safar SV, Gomes FC, Sousa FM, Morais PB, Broetto L, Leoncini O, Ribeiro JR, Fungsin B, Takashima M, Nakase T, Lee CF, Vainstein MH, Fell JW, Scorzett G, Vishniac HS, Rosa CA, Valente P. Hannaella pagnoccae sp. nov., a tremellaceous yeast species isolated from plants and soil. *Int J Syst Evol Microbiol*. 2014;64:1970-7.
35. Sidhu GS. Gibberella spp., pathogens of many crop species. In: Sidhu GS, editor. *Genetics of Plant Pathogenic Fungi*. Advances in plant Pathology. Volume 6. London: Academic Press Limited; 1988. p. 159-67.
36. Howard PC, Eppley RM, Stack ME, Warbritton A, Voss KA, Lorentzen RJ, Kovach RM, Bucci TJ. Fumonisin b1 carcinogenicity in a two-year feeding study using F344 rats and B6C3F1 mice. *Environ Health Perspect*. 2001;109 Suppl 2:277-82.
37. Prohic A, Jovovic Sadikovic T, Krupalija-Fazlic M, Kuskunovic-Vlahovljak S. Malassezia species in healthy skin and in dermatological conditions. *Int J Dermatol*. 2016;55:494-504.
38. Bladt TT, Frisvad JC, Knudsen PB, Larsen TO. Anticancer and antifungal compounds from Aspergillus, Penicillium and other filamentous fungi. *Molecules*. 2013;18:11338-76.

39. Lou J, Fu L, Peng Y, Zhou L. Metabolites from *Alternaria* fungi and their bioactivities. *Molecules*. 2013;18:5891-935.
40. Nidhiya KA, Satya E, Nitya MR, Saranya A, Suganthi R. In vitro anticancer potential of statin from *Aspergillus tamarii* grd119. *Asian J Pharm Clin Res*. 2015;8:147-53.
41. Raj KG, Sambantham S, Manikanadan R, Arulvasu C, Pandi M. Fungal taxol extracted from *Cladosporium oxysporum* induces apoptosis in T47D human breast cancer cell line. *Asian Pac J Cancer Prev*. 2014;15:6627-32.
42. Page RC, Eke PI. Case definitions for use in population-based surveillance of periodontitis. *J Periodontol*. 2007;78:1387-99.

Supplementary Information

Supplementary Table 1. Abundances (Sheet 1) and detection frequencies (Sheet 2) of phyla detected in each of the samples and across the study groups.

Supplementary Table 2. Abundances (Sheet 1) and detection frequencies (Sheet 2) of genera detected in each of the samples and across the study groups.

Supplementary Table 3. Abundances (Sheet 1) and detection frequencies (Sheet 2) of species detected in each of the samples and across the study groups.

Supplementary Table 4. List of taxa exclusively identified in either group at prevalence $\geq 10\%$.

Figure Legends

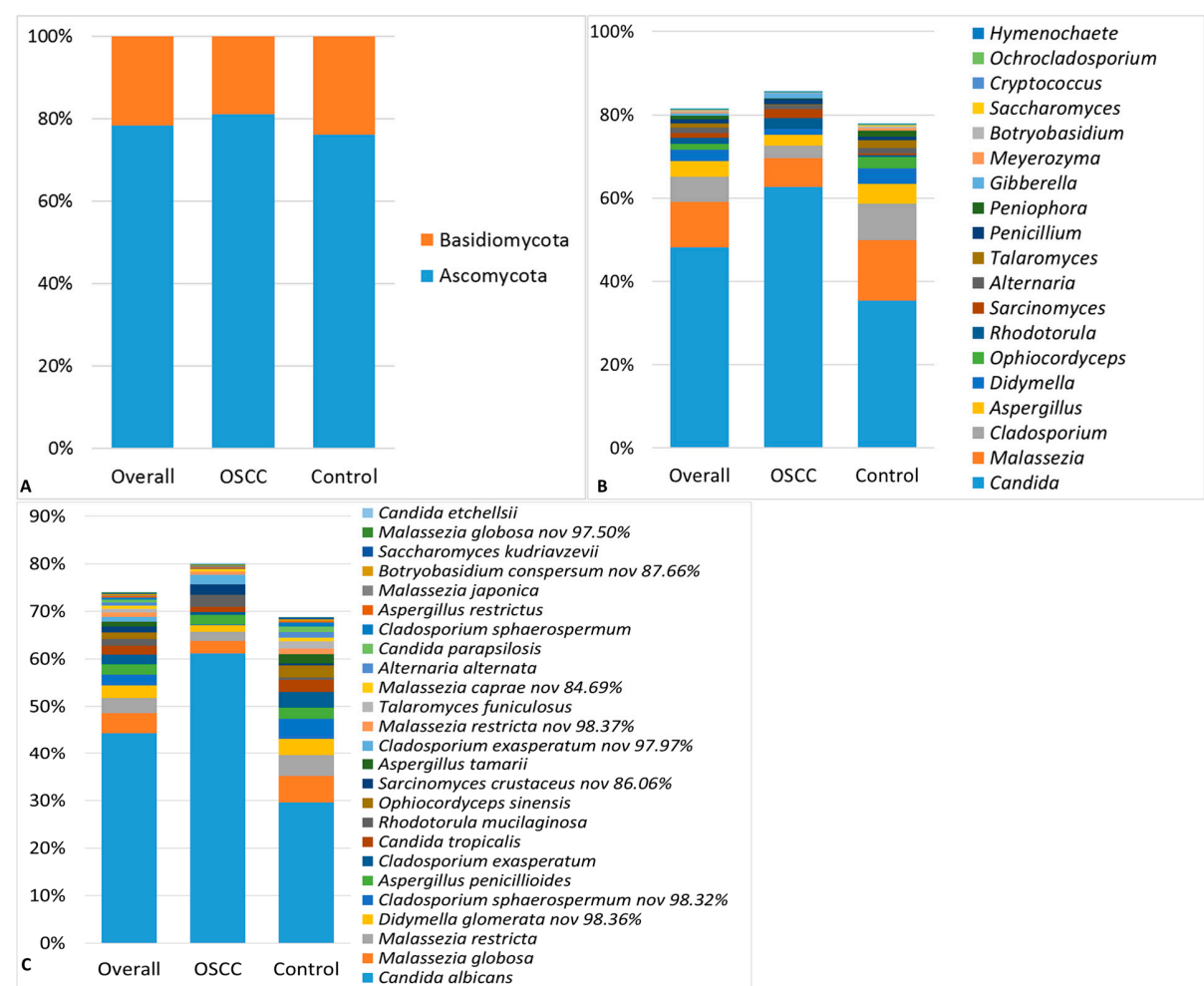


Figure 1. Mycobiome profile. Average abundances of the two phyla identified (A) as well as the genera (B) and species (C) detected in $\geq 15\%$ of the samples.

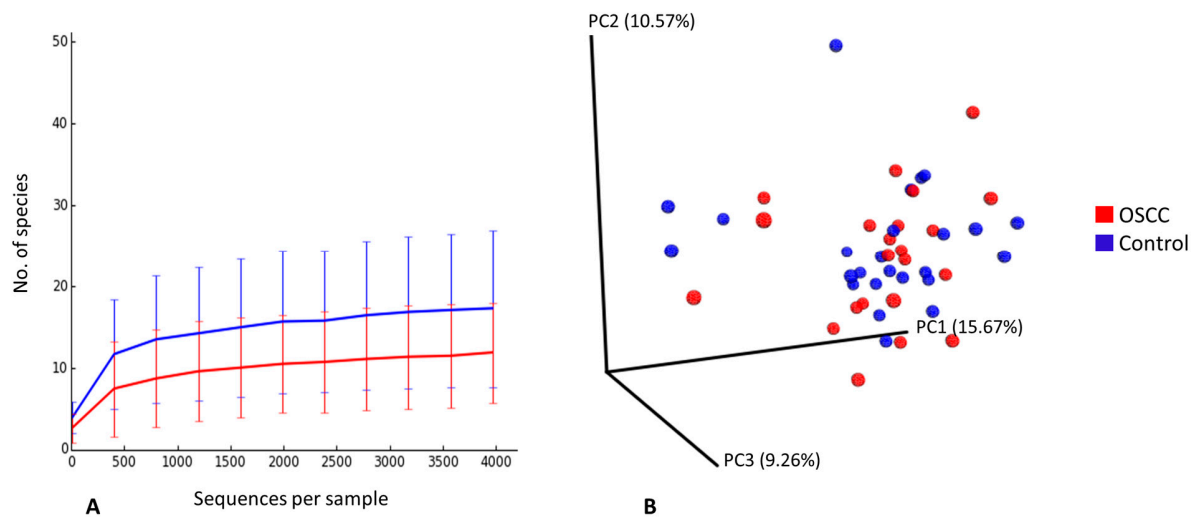


Figure 2. Rarefaction and β -diversity. A. Rarefaction curves showing the number of observed species as a function of sequencing depth. B. Non-clustering of the study subjects by PCoA (weighted Unifrac).

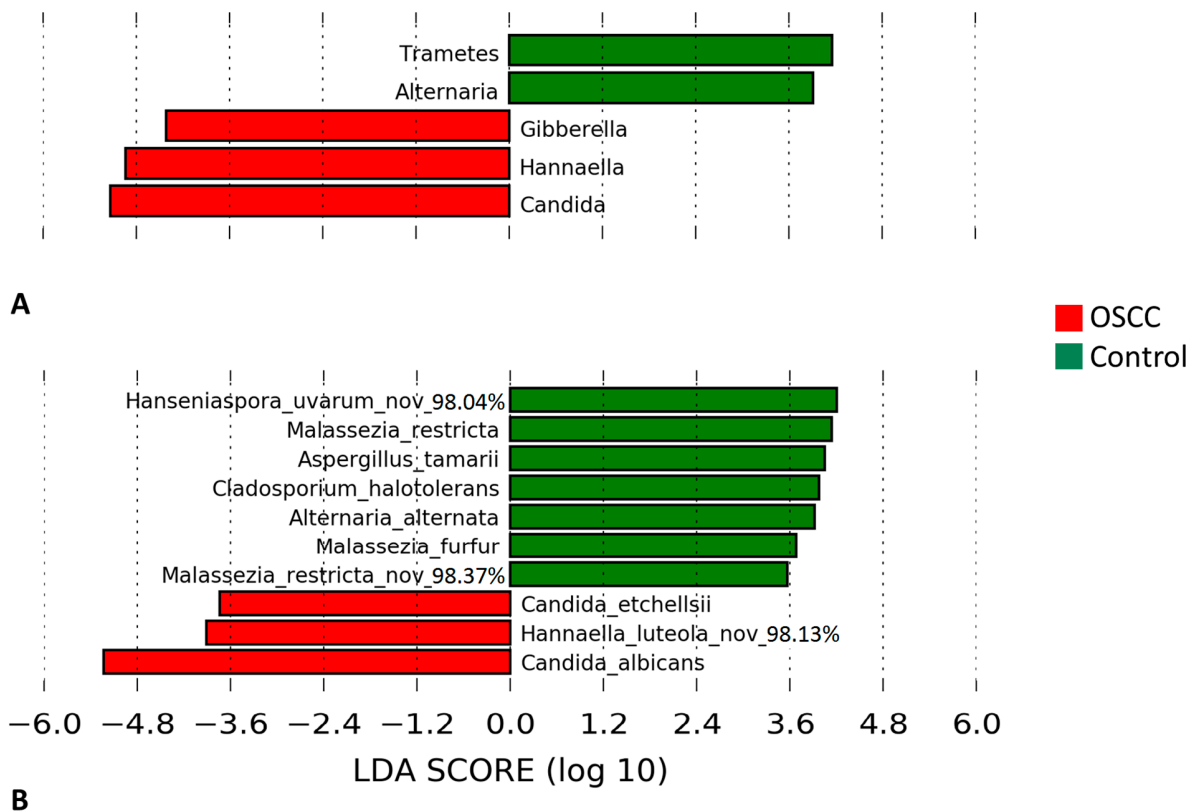


Figure 3. Differentially abundant taxa. Linear Discriminant Analysis Effect Size (LEfSe) analysis showing genera (A) and species (B) that were significantly differentially abundant between the cases and controls (LDA score ≥ 3).

Table 1. Clinical characteristics of the study subjects

Variable	Cases (n=22)	Controls(n=25)
Age (mean±SD) *	61.00±9.5	50.58 ±13.5
% males	100	100
Betel quid: No. (%) ¶		
Never	0 (0.0)	4 (16)
Past	3(13.6)	2 (32)
Sometimes	1 (4.6)	7 (28)
Daily	18 (81.9)	12 (48)
Smoking: No. (%) ¶		
Never	4 (18.2)	8 (32)
Past	5 (22.7)	7 (28)
Sometimes	4 (18.2)	4 (16)
Daily	9 (40.9)	6 (24)
Alcohol use: No. (%) ¶		
Never	0 (0.0)	4 (16)
Past	5 (22.7)	2 (8)
Sometimes	4 (18.2)	15 (60)
Daily	13 (59.1)	4 (16)
Oral hygiene: No. (%) ¶		
Bad	6 (27.3)	3 (12)
Fair	14 (63.6)	10 (40)
Good	2 (9.1)	12 (48)
Missing teeth (mean±SD) *	14.73±10.01	3.85±4.82
Periodontal status: No. (%) ¶§		
Severe Periodontitis	8 (36.4)	2 (3)
Moderate Periodontitis	11 (50.0)	6 (24)
No or Mild Periodontitis	3 (13.6)	17 (68)
Site affected: No. (%)		
Tongue	14 (63.6)	19 (76)
Buccal mucosa	8 (36.4)	6 (24)
Histopathology: No. (%)		
Well differentiated SCC	11 (50)	19 (76)
Moderately differentiated SCC	11 (50)	6 (24)
Fibro-epithelial polyp	0 (0)	25 (100)

* Statistically significant: $P \leq 0.05$; t-test¶ Statistically significant: $P \leq 0.05$; Chi-squared test

§ Classification made according to Page and Eke [42]

Table 2. Species richness, α -diversity and coverage (mean \pm SE) calculated from the rarefied biom.

Group	Observed richness	Chao1	Shannon index	Good's coverage
OSCC	11.7 \pm 6.4	13.6 \pm 7.0	1.5 \pm 0.9	0.999 \pm 0.000
FEP	17.7 \pm 9.8	19.9 \pm 12.2	2.1 \pm 1.2	0.999 \pm 0.001



© 2017 by the authors. Licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).