Check for updates

RESEARCH ARTICLE

Mycobiome analysis in fungal infected formalin-fixed and paraffin-embedded tissues for identification of pathogenic fungi: a pilot study [version 1; peer review: 1 approved, 1 approved with reservations]

Taebum Lee¹, Hee Young Na², Sun-ju Byeon³ , Kyoung-Mee Kim¹,
Hey Seung Lee², Sung-Hye Park⁴, Ji-Young Choe⁵, Kyoung Chan Choi⁶

¹Pathology and Translational Genomics, Samsung Medical Center, Seoul, South Korea

²Pathology, Seoul National University Bundang Hospital, Seongnam-si, South Korea

³Pathology, Hallym University Dongtan Sacred Heart Hospital, Hwaseong-si, South Korea

⁴Pathology, Seoul National University Hospital, Seoul, South Korea

⁵Pathology, Hallym University Sacred Heart Hospital, Anyang-si, South Korea

⁶Pathology, Chuncheon Sacred Heart Hospital, Chuncheon-si, South Korea

V1 First published: 22 Jul 2020, 9:758
<https://doi.org/10.12688/f1000research.25126.1>

Latest published: 22 Jul 2020, 9:758
<https://doi.org/10.12688/f1000research.25126.1>

Abstract

Background: Fungal organisms are frequently observed in surgical pathological diagnosis. In order to more accurately identify fungi in formalin-fixed and paraffin-embedded (FFPE) tissues, it is necessary to use genomic information. The purpose of our pilot study is to identify the factors to be considered for the identification of pathogenic fungi using mycobiome analysis in FFPE tissues.

Methods: We selected 49 cases in five hospitals. In each case, FFPE tissue was cut into 50 µm and DNA was extracted. Multiplex PCR with four primers (ITS1, ITS2, ITS3 and ITS4) was performed. Multiplex sequencing was performed using a MinION device according to the manufacturer's protocol. Sequences of each case were searched using BLASTN with an ITS database from NCBI RefSeq Targeted Loci Project with default parameters.

Results: A total of 2,526 DNA sequences were sequenced. We were able to identify 342 fungal sequences in 24 (49.0%, 24/49) cases. The median number of detected fungal sequences per case was 3 (1Q: 1 and 3Q: 14.25). Of the fungal DNA sequences, 215 (62.87%) contained the entire region of ITS1 or ITS2. The remaining 127 fungal DNA sequences were identified as fungi using a partial sequence of ITS1, ITS2, 5.8S, LSU or SSU.

Conclusion: In conclusion, we have identified the possibility of finding pathogenic fungi through mycobiome analysis in fungal infected FFPE tissues using nanopore sequencing. However, we have also found

Open Peer Review

Approval Status

	1	2
version 1
22 Jul 2020	view	view

1. **Wissam Faour** , Lebanese American University, Byblos, Lebanon

2. **Chengqiang Wang**, Shandong Agricultural University, Tai'an, China

Any reports and responses or comments on the article can be found at the end of the article.

several limitations to be solved for further studies. If we develop a method to characterize pathogenic fungi in FFPE tissues in a follow-up study, we think it will help patients to use appropriate antifungal agents.

Keywords

Nanopores, Fungi, Genomics, Mycobiomes

Corresponding author: Sun-ju Byeon (byeon.sunju@welovedoctor.com)

Author roles: **Lee T:** Data Curation, Formal Analysis, Writing – Original Draft Preparation; **Na HY:** Data Curation, Formal Analysis, Writing – Original Draft Preparation; **Byeon Sj:** Conceptualization, Project Administration, Validation, Writing – Review & Editing; **Kim KM:** Data Curation; **Lee HS:** Data Curation; **Park SH:** Data Curation; **Choe JY:** Data Curation; **Choi KC:** Data Curation, Resources

Competing interests: No competing interests were disclosed.

Grant information: This research was supported by Hallym University Research Fund 2018 [HURF-2018-30].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Lee T *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lee T, Na HY, Byeon Sj *et al.* **Mycobiome analysis in fungal infected formalin-fixed and paraffin-embedded tissues for identification of pathogenic fungi: a pilot study [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2020, 9:758 <https://doi.org/10.12688/f1000research.25126.1>

First published: 22 Jul 2020, 9:758 <https://doi.org/10.12688/f1000research.25126.1>

Introduction

Fungal organisms are frequently observed in surgical pathological diagnosis. Recent developments in technology have enabled the identification of fungi using a variety of methods¹. These novel methods are based on analyzing fungal genomes or proteomes using fresh tissues^{2,3}. However, these methods are difficult to apply in formalin-fixed and paraffin-embedded (FFPE) tissues used in surgical pathology. Sanger sequencing and immunohistochemistry staining have been used in FFPE tissue, but most medical institutions only identify fungi based on morphological findings⁴⁻⁹. The identification of fungi only by morphological findings can lead to inadequate treatment for patients due to misdiagnosis, which can often result in fatal consequences¹⁰. As different antifungal agents are preferentially used at the initial infection stage depending on the fungus, it is necessary to identify the exact fungi present through testing methods in addition to the morphological findings¹¹.

In order to more accurately identify fungi in FFPE tissues, it is essential to use genomic information. DNA markers that could be used to identify fungi include the internal transcribed spacer (ITS) region, small subunit (nrSSU-18S), large subunit (nrLSU-26S or 28S), elongation factor 1- α (EF1 α), and the largest (RPB1) and second largest (RPB2) subunits of RNA polymerase^{1,2}. Among these markers, ITS could be relatively easily and effectively used for fungal identification, and the database of the ITS regions of fungi is available¹²⁻¹⁴. In general, sequencing equipment is classified into first- (Sanger sequencing), second- (massively parallel sequencing), and third-generation (real-time and single molecule sequencing) equipment according to key analytical methods¹⁵. Third-generation sequencing technology is characterized by direct sequencing of nucleotides without PCR amplification. Oxford Nanopore Technologies (ONT) introduced several sequencing devices using nanopore sequencing technology, which measures the change in current that occurs when a nucleotide sequence passes through a narrow channel¹⁶. There is an advantage in identifying each DNA sequence without PCR amplification, and it is expected that the DNA can be effectively detected in spite of DNA degradation during FFPE tissue preparation and storage.

Extracting only fungal DNA from the fungal infection site of FFPE tissue is very difficult. Therefore, we decided to use the mycobiome analysis method. However, no studies have attempted to analyze the mycobiome in FFPE tissue. Therefore, this study was conducted as a pilot study on the development of a method for finding pathogenic fungi using mycobiome analysis in fungal infected FFPE tissues.

Methods

Sample collection

The cases were extracted from the pathological examination reports of five medical institutions. We first extracted pathology reports that mentioned the presence of fungi. We then selected typical cases for use as positive controls, cases reported to be difficult to differentiate (briefly, fungi with branched-hyphae, such as *Aspergillus* species and *Mucor* species, are often difficult to distinguish morphologically), and cases with additional

information related to fungal identification on the pathology report (i.e., identification using culture or sequencing). Through this process, we finally selected 49 cases (case 3-02 and 5-10 were able to confirm the fungal identification results using sequencing). Cases included 21 lung, eight paranasal sinus, seven gastrointestinal tract, five orbit, two skin and mouth, one adrenal gland, one bone, one gum and one liver samples. The average storage period of FFPE tissue was 4.4 years. This study was exempted from obtaining informed consent by the Institutional Review Board of Hallym University Dongtan Sacred Heart Hospital (NON2018-005).

DNA extraction, PCR amplification, sequencing and base calling

In each case, FFPE tissue was cut into 50 μ m (5 μ m $\times$ 10) sections and collected in a 2.0 ml conical tube. DNA was extracted using the ReliaPrep FFPE gDNA Miniprep System (Catalogue number: A2351, Promega, Madison, Wisconsin, USA) according to the [manufacturer's protocol](#). The DNA extracted from all cases and 10 μ l DNA was loaded onto 1% agarose gel (Certified Molecular Biology Agarose, #1613101, Bio-Rad Laboratories, Hercules, California, USA, Dyne LoadingSTAR with 100bp DNA Ladder, #A751, Dynebio, Seongnam, Korea, Mupid-One, #AD160, Takara Bio, California, USA) with 100V for 30 minutes running to confirm the presence of DNA. Four universal primers (ITS1, 5'-TCCGTAGGTGAACCTGCGG-3'; ITS2, 5'-GCTGCGTTCTTCATCGATGC-3'; ITS3, 5'-GCATCG ATGAAGAACGCAGC-3'; and ITS4, 5'-TCCTCCGCTTAT TGATATGC-3') were ordered from Macrogen (Seoul, Korea) with polyacrylamide gel electrophoresis purification¹⁷. Multiplex PCR with four primers using SimpliAmp Thermal Cycler (ThermoFisher Scientific, Waltham, Massachusetts, USA) and Taq polymerase (New England Biolabs, M0287S, Ipswich, Massachusetts, USA) was performed as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 1 min with a final extension step of 72°C for 8 min and cooling at 4°C¹. Sequencing was performed using the FLO-MIN106 flow cell (ONT, Oxford Science Park, UK), ligation sequencing kit 1D (SQK-LSK109, ONT) and PCR bar-coding expansion 1-12 (EXP-PBC001, ONT) with a MinION (ONT) device in accordance with the manufacturer's protocol (protocol version: [PBGE12_9066_v109_revC_23May2018](#)). Multiplex sequencing was performed for 10-12 cases at a time (83-100 ng DNA per sample). The base calling was carried out according to ONT's recommendations without parameter changes using MinION Software Release 18.12.6.

Data analysis

We used BLAST+ 2.10.0 on a local PC (Ubuntu 19.10) for analysis¹⁸. In brief, we downloaded the ITS database file from NCBI RefSeq Targeted Loci Project (accession number: [PRJNA177353](#), last modified Mar 3, 2020) and then converted it using "makeblastdb" of BLAST+ 2.10.0. This database file contained the full or partial ITS sequence for 11,133 genera of fungi. The RU7 genomic reference sequences of *Pneumocystis jirovecii* (accession number: [GCF_001477535.1](#)), which are not found in this database, were downloaded and compared. The sequence of *Actinomyces israelii*, a bacterium that looks

morphologically similar to that of a filamentous fungus, was downloaded (*Actinomyces israelii* DSM 43320, whole genome shotgun sequencing project; [NZ_JONS000000000.1](#)) and compared. We confirmed the BLASTN search using GRCh38 (accession number: [GCF_000001405.26](#)) to determine if the detected DNA corresponds to the human genome.

We converted FASTQ files to FASTA format using “seqtk” on Github. Sequences of each case were searched using BLASTN using the ITS database file without default parameter modification. When the sequence matched in the ITS database, the result with the highest bit score value was selected. We selected five results in order of highest bit score when five or more fungal DNAs were detected in one case.

We downloaded the GenBank files of all fungi contained in the database and separated the base sequences of ITS1, 5.8S and ITS2 for each fungus. When fungi were detected, they were classified into three categories (“entire”, “partial”, and “none”) according to the relationship between the detected fungal DNA and ITS1/ITS2. “Entire” means that all ITS1 or ITS2 base information is used for fungal identification, “partial” means that identification is based on a part of ITS1 or ITS2 and its surrounding sequence, and “none” means that base information other than ITS1 or ITS2 is used.

An earlier version of this article can be found on bioRxiv (<https://doi.org/10.1101/2020.04.19.045856>).

Results

A total of 2,526 DNA sequences were sequenced¹⁹. No DNA was sequenced in three cases. Of the remaining 46 cases, the sequenced DNA of 22 cases were not found in the ITS database. We were able to identify 342 fungal sequences in 24 (49.0%, 24/49) cases. The number of DNA sequences identified per case are summarized in [Figure 1](#). The median number of the detected fungal DNA sequences per case was 3 (1Q: 1 and 3Q: 14.25). The clinicopathological information and the number of fungal species detected per case are summarized in [Table 1](#). The remaining cases, in which fungal DNA was not sequenced, are listed in [Table 2](#). Detailed information related to the BLAST program is summarized in Supplementary Table 1 (see *Extended data*)¹⁹. The relationship between the detected fungal DNA and the ITS region is summarized in [Table 3](#) and visualized in [Figure 2](#). Most (215, 62.87%) fungal DNA sequences contained the entire region of ITS1 or ITS2. The remaining 127 fungal DNA sequences were identified as fungi using a partial sequence of ITS1, ITS2, 5.8S, LSU or SSU.

The mean FFPE block storage period of fungal and no fungal cases was 4.2 years and 4.8 years, respectively ($p = 0.410$). The

Figure 1. Summary of the number of DNA sequences identified per case.

Table 1. The basic clinicopathological information of fungal DNA confirmed cases (underlined when a fungus identical to pathological diagnosis is detected).

Case	Sex/ Age	Location	Pathologic diagnosis	Sequencing reads (fungal DNA)	Number of fungal DNA reads	Species	Storage years	DNA concentration (ng/μl)	Note
1-01	M/38	Nasal cavity	Actinomyces	3 (3)	2	<u>Candida</u>	7	7.55	
1-02	F/67	Bone	Actinomyces	39 (21)	5	<u>Aspergillus</u>	5	40.8	
					4	<u>Acremonium</u>			
					2	<u>Aspergillus</u>			
					1	<u>Candida</u>			
1-03	F/80	Nasal cavity	Mucormycosis	16 (4)	3	<u>Cladosporium, Fusarium, Ganoderma, Moniliella, Mycoleptodonoides, Ochrocladosporium, Peniophora, Sistotremastrum, Wallenia, Xylodon</u>	4	60.75	
					1	<u>Aspergillus</u>			
1-04	F/80	Orbit	Mucormycosis	40 (11)	2	<u>Acremonium</u>	4	15.4	
					1	<u>Candida, Ganoderma</u>			
1-06	F/80	Orbit	Mucormycosis	6 (1)	1	<u>Aspergillus, Favolus, Miliopium, Phaeophlebiopsis, Rhizopus, Theleporus, Xylodon</u>	4	131.1	
1-07	F/78	Esophagus	Candidiasis	55 (33)	24	<u>Candida</u>	4	88.5	
					2	<u>Xylodon</u>			
1-08	M/61	Gingiva	Actinomyces	107 (85)	16	<u>Acremonium, Aspergillus, Collembolespora, Inocybe, Otidea, Sarcodon, Trichomerium</u>	4	7.8	
					11	<u>Alternaria</u>			
					6	<u>Fusarium</u>			
					4	<u>Aspergillus</u>			
					3	<u>Gladosporium</u>			
					2	<u>Neofusicoccum, Sporormurispora</u>			
					1	<u>Candida, Penicillium, Platystomum, Sarcodon, Stammerella, Torulaspora</u>			
					1	<u>Acaulospora, Aleurodiscus, Antrodia, Ascochyta, Auriculoscypa, Biappendiculipora, Butleria, Cerrina, Cookeina, Corynespora, Hydnothlebia, Hypoxylon, Inocybe, Kurtzmanomyces, Lysurus, Megasporia, Neobulgaria, Papillotrema, Paracradaphialophora, Phylloporia, Pleospora, Pluteus, Pseudofusicoccum, Pseudophiostoma, Sclerotium, Sterigmatomyces, Trichoderma, Tylophila, Udeniomyces, Wickerhamiella</u>			

Case	Sex/ Age	Location	Pathologic diagnosis	Sequencing reads (fungal DNA)	Number of fungal DNA reads	Species	Storage years	DNA concentration (ng/ul)	Note
1-09	M/81	Esophagus	Candidiasis	66 (30)	5	<i>Fantasmomyces</i>	2	7.85	
					4	<i>Aspergillus</i>			
					2	<i>Camarographium</i> , <i>Candida</i> , <i>Phaeophielebiopsis</i>			
					1	<i>Cercospora</i> , <i>Cladosporium</i> , <i>Donkioporiella</i> , <i>Dothiora</i> , <i>Elaphomyces</i> , <i>Laetisaria</i> , <i>Miliospium</i> , <i>Penicillium</i> , <i>Periconia</i> , <i>Pleiospora</i> , <i>Russula</i> , <i>Sigarispora</i> , <i>Trametes</i> , <i>Truncospora</i> , <i>Veluticeps</i>			
1-10	M/70	Esophagus	Candidiasis	123 (111)	103	<i>Candida</i>	2	4.45	
					5	<i>Otidea</i>			
					1	<i>Alternaria</i> , <i>Cladosporium</i> , <i>Lactifluus</i>			
2-03	M/42	Esophagus	Unknown	1 (1)	1	<i>Acremonium</i>	3	24.8	
2-07	M/56	Skin	Mucormycosis	1 (1)	1	<i>Candida</i>	2	85.25	
2-08	F/57	Lung	Unknown	5 (1)	1	<i>Starmerella</i>	1	10.65	
2-09	F/50	Pharynx	Unknown	4 (1)	1	<i>Cladosporium</i>	9	5.25	
3-02	M/47	Lung	Blastomycosis	6 (1)	1	<i>Candida</i>	8	285.8	
3-03	M/69	Lung	Cryptococcosis	104 (5)	3	<i>Candida</i>	2	192.15	
					1	<i>Acremonium</i> , <i>Aspergillus</i>			
3-04	F/69	Skin	Unknown	147 (14)	9	<i>Candida</i>	2	28.55	
					1	<i>Aspergillus</i> , <i>Fusarium</i> , <i>Penicillium</i> , <i>Saccharomyces</i> , <i>Violella</i>			
3-05	F/78	Oral cavity	Mucormycosis & Candidiasis	19 (2)	2	<i>Candida</i>	8	11.4	
3-07	M/68	Nasal cavity	Mucormycosis	3 (1)	1	<i>Candida</i>	7	164	
3-10	M/72	Lung	Cryptococcosis	69 (6)	3	<i>Candida</i>	2	98.6	
					1	<i>Aspergillus</i> , <i>Cryptococcus</i> , <i>Fusarium</i>			
3-11	F/55	Lung	Mucormycosis	36 (3)	1	<i>Alternaria</i> , <i>Candida</i> , <i>Rhizopus</i>	4	252.7	
4-03	F/77	Orbit	Mucormycosis	38 (1)	1	<i>Rhizopus</i>	3	25.05	
4-07	M/20	Lung	Pneumocystis	133 (1)	1	<i>Rhizopus</i>	6	94.55	
4-12	M/64	Lung	Pneumocystis	328 (4)	2	<i>Candida</i>	7	30.45	Pneumocystis jirovecii with bit score 73.1
					1	<i>Aspergillus</i> , <i>Cladosporium</i>			
5-07	F/59	Lung	Cryptococcosis	18 (1)	1	<i>Candida</i>	1	34.7	

Table 2. The basic clinicopathological information of no-fungal DNA detected cases (sorted by pathologic diagnosis).

Case	Sex/Age	Location	Pathology diagnosis	Number of sequences	Storage years	DNA concentration (ng/μl)
2-01	M/51	Stomach	Actinomyces	0 (No DNA)	5	24.55
2-02	F/40	Eye	Actinomyces	1	3	12.4
2-05	F/42	Paranasal sinus	Aspergillosis	1	2	14.7
2-06	F/58	Paranasal sinus	Aspergillosis	0 (No DNA)	2	8.9
3-01	M/60	Lung	Aspergillosis vs. Mucormycosis	10	8	540
4-10	M/73	Paranasal sinus	Aspergillosis vs. Mucormycosis	175	6	122.75
5-03	M/88	Paranasal sinus	Aspergillosis vs. Mucormycosis	21	3	Not applicable
5-04	M/64	Lung	Aspergillosis vs. Mucormycosis	30	2	10
2-04	M/52	Stomach	Candida	1	3	36.25
3-08	M/63	Lung	Cryptococcus	5	9	269.9
3-09	F/51	Lung	Cryptococcus	6	6	59.45
5-05	M/61	Lung	Cryptococcus	84	2	25.45
5-06	F/60	Lung	Cryptococcus	3	1	36.15
5-10	F/11	Liver	Cryptococcus	34	12	212.6
5-01	M/58	Terminal ileum	Histoplasmosis	23	5	46.6
5-02	M/58	Adrenal gland	Histoplasmosis	71	5	14.2
1-05	F/79	Paranasal sinus	Mucormycosis	0 (No DNA)	4	92.9
3-06	M/70	Orbit	Mucormycosis	1	7	63.95
3-12	F/14	Lung	Mucormycosis	20	4	130
4-01	M/44	Lung	Mucormycosis	32	3	71.4
4-02	M/56	Lung	Mucormycosis	259	3	123.25
4-04	F/50	Lung	Mucormycosis	143	1	156
4-05	F/4	Lung	Pneumocystis	147	9	288.85
4-06	M/5	Lung	Pneumocystis	13	8	68.85
4-09	M/48	Lung	Pneumocystis	78	8	76.5

Table 3. Classification according to the degree to which the base sequence used to identify fungi matches the ITS1 or ITS2 region (the number in parentheses is percent).

		ITS2			
		Entire	Partial	None	Total
ITS1	Entire	9 (2.63)	1 (0.29)	103 (30.12)	113 (33.04)
	Partial	4 (1.17)	1 (0.29)	58 (16.96)	63 (18.42)
	None	98 (28.65)	51 (14.91)	17 (4.97)	166 (48.54)
	Total	111 (32.46)	53 (15.50)	178 (52.05)	342 (100)

Figure 2. Visualization of nucleotide sequences used for fungal identification according to the degree of ITS1 and ITS2 sequence matching (the nucleotide sequences of the reference database were aligned with the starting base of the ITS2 region).

DNA concentrations measured by NanoDrop after DNA extraction in the groups with and without fungal DNA detection were 71.2 ng/μl and 104.4 ng/μl, respectively, with no statistical significance ($p=0.272$). The fungal DNA detection rate was not statistically significant ($p=0.376$).

Representative microscopic images of some cases with inconsistent pathologic diagnosis and mycobiome analysis are summarized in Figure 3 (original images are provided as *Underlying data*)²⁰. *Aspergillus* and *Candida* species were detected by sequencing in case 1-01 (Figure 3A). Compared to the *Aspergillus* species commonly found in nasal cavities, thinner hyphae were observed, which may have been misleadingly similar to sulfur granules of the pathologically diagnosed *Actinomyces*. *Aspergillus* and *Acremonium* species were detected by sequencing in case 1-03 (Figure 3B). Contrary to case 1-01, it is thought to have been misdiagnosed as *Mucor*

species because its hyphae appeared slightly wider than *Aspergillus* species. Case 1-08 (Figure 3C) was also misdiagnosed as *Actinomyces*, in what is thought to be a similar manner to case 1-01. *Alternaria*, *Fusarium* and *Aspergillus* species were detected with numerous sequencing reads. In case 2-08 (Figure 3D), yeast-form fungi were observed in alveolar macrophages using a Gomori methenamine-silver stain and *Starmerella cellae* was identified by sequencing. *Starmerella cellae* is a relatively recently identified ovoid to ellipsoidal fungus²¹. Case 2-09 (Figure 3E) is a fungus found in the pharynx, which shows morphological findings different from those of *Candida*, *Aspergillus*, and *Mucor*. In other words, yeast-form or short branching-type fungal nuclei were found, like in *Pneumocystis jiroveci*. *Cladosporium coloradense* was identified by sequencing. Case 3-04 (Figure 3F) is a yeast-form fungus found in subcutaneous tissue and *Candida glabrata* was identified by numerous sequencing reads.

Figure 3. Representative microscopic images with inconsistent pathological and molecular diagnoses (magnification x600). A) case 1-01, **B)** case 1-03, **C)** case 1-08, **D)** case 2-08, **E)** case 2-09 and **F)** case 3-04.

All three cases diagnosed with *Actinomyces* were found to have fungi (cases 1-01, 1-02 and 1-08). Case 3-02 was had a pathological diagnosis of *Blastomyces dermatitidis* by sequencing analysis (summary of sequencing results: 575 nucleotide sequences of [U18364.1](#) were identical without gap opening

and 549 nucleotide sequences of [EF592163.1](#) were identical without gap opening). One strand of fungal DNA was identified from the DNA extracted from the FFPE tissue, and the BLASTN analysis showed the highest probability match was *Candida africana* with a bit score of 468. In this case, the bit score was 154

based on the separate sequence of *Blastomyces dermatitidis*. Among three cases of pathologically diagnosed cryptococcosis, one case (case 3-10) was found to have a nucleotide sequence of *Cryptococcus neoformans* with a bit score of 296. In case 4-12, the nucleotide sequence of *Candida africana* with a bit score of 438 was suggested, with the possibility of *Pneumocystis jiroveci* with a bit score of 73.1. Four cases of pathologically unidentifiable fungi (cases 2-03, 2-08, 2-09 and 3-04) were found to most likely be *Acremonium acutatum*, *Starmerella cellae*, *Cladosporium coloradense*, and *Candida glabrata*, respectively.

Discussion

We performed mycobiome analysis on fungal infected FFPE tissues using nanopore sequencing. The detected fungal DNA occupied approximately one-third of the ITS1 entire region, the ITS2 entire region, and other regions. The advantages of nanopore sequencing compared with Sanger sequencing are as follows. First, unlike Sanger's method, which requires a lot of DNA, nanopore sequencing can be performed with a small amount of DNA (in theory, even with one strand). Second, the nanopore sequencing method can sequence DNA separately, even if the sample contains a variety of lengths of DNA²². Third, nanopore sequencing equipment (i.e., MinION) can be operated at a lower cost (about \$1,000) than Sanger sequencing equipment. This low initial cost is a critical factor in the introduction of equipment in small pathology laboratories. Compared to second-generation sequencing equipment, nanopore sequencing has the advantage of sequencing damaged DNA because there is no PCR amplification process in the sequencing process itself. In this study, about one third of the fungal DNA was not an ITS1 entire match or an ITS2 entire match. Therefore, these fungi can be effectively detected using nanopore sequencing.

It is necessary to find pathogenic fungi among various fungi detected by mycobiome analysis. It may be assumed that dominant fungi are associated with the disease at the site of fungal infection. In cases 1-07 and 1-10, diagnosed as esophageal candidiasis, *Candida* species comprised 72% and 93% of the fungal DNA, respectively. In such cases, it would be clear to diagnose that the infection is caused by *Candida* species. In case 1-09, however, only 7% of the fungal DNA was *Candida* species. Similarly, *Rhizopus* species were detected in *Mucor* infections, but as only a small fraction of the fungal DNA. Considering these cases, it is considered desirable to include the process of detecting normal flora in adjacent uninfected sites and selecting pathogenic fungi associated with infection.

We performed multiplex PCR to increase fungal DNA concentration. However, 2,184 non-fungal DNA sequences accounted for 86.46% of the total DNA. Of the 2,184 DNA sequences not identified as fungi, only 382 (17.50%) were identified in the human genome (GRCh38). Because ribosomal DNA is present in all living organisms, including fungi, humans, and bacteria, it may be difficult to amplify only the fungal ITS region. Nevertheless, we were only able to detect fungal DNA

in 49.0% of cases. This detection rate is not high enough to be applied in actual clinical situation and needs improvement. There was no difference in the storage period or DNA quality of the FFPE block between cases where fungal DNA was detected and cases that were not detected. We suggest the following about the causes of low fungal DNA detection rates. The first and most important is a low sequencing output. The third-generation sequencing technique is known to have less sequencing output than the second-generation sequencing equipment. Therefore, increasing the purity of the DNA to be sequenced is important for research. Since we used multiplex sequencing, we were forced to reduce the amount of DNA we analyzed per sample. We believe that the lack of sufficient DNA sequencing in each case was the main cause of the failure to detect fungal DNA. The second is the lack of fungal DNA evaluation methods. Unlike human cells, in which cell viability (closely related to DNA quality) could be predicted by evaluation of a haematoxylin and eosin stained slide, how to evaluate the viability of fungi in a haematoxylin and eosin stain is not known. Many studies have shown that inflammation and hypoxia are closely related, and hypoxia induced by inflammation caused by fungal infections may result in damage to fungal DNA and consequently affect the sequencing²³. Third, the aforementioned ribosomal DNA is present in almost all living things. We believe this could be overcome by using microdissection, which extracts DNA from fungi as much as possible. Fourth, there is a lack of high-quality databases. Even in the database of NCBI RefSeq Targeted Loci Project, there is only one species (*murina*) in the *Pneumocystis* genus. In other words, to diagnose *Pneumocystis jiroveci*, a separate database should be constructed. In addition, when we analyzed the GenBank file, there were 5,602 ITS1 full sequences and 7,552 ITS2 full sequences in 11,133 genera. Therefore, one well-known fungal ITS DB is not sufficient for identification of fungi, and several DBs must be synthesized to analyze the results. Fifth, it is known that tumor biopsy specimens of 50µm thickness can be used to sufficiently perform next-generation sequencing for diagnostic purposes²⁴. Our study was also performed with a 50µm thickness. However, if thicker tissues (e.g., 100µm thick) were used, the fungal detection rate could be increased.

Candida albicans is the most commonly isolated *Candida* species in the clinical setting. Of the 162 *Candida* species, 141 *Candida africana* (87.04%) and eight *Candida albicans* (4.94%) sequences were detected. These two *Candida* species are similar in ITS sequences. Comparing the ITS sequences (ITS1+5.8S+ITS2) of *Candida africana* (NR_138276.1, 447bp) with *Candida albicans* (NR_125332.1, 446bp), there is 99.192% identity. There is a 1 bp gap in ITS1 in *Candida albicans*. Two mismatched nucleotides are located in ITS2, where the DNA base "C" in *Candida albicans* is "T" in *Candida africana*. It is generally known that nanopore sequencing has a slightly higher error rate than second-generation NGS sequencing²⁵. DNA extracted from FFPE tissue is known to have a higher C:T conversion than DNA extracted from fresh frozen tissue²⁶. Because of these, it is difficult to use our method to precisely identify fungi, especially at the species level.

In summary, we have identified the possibility of finding pathogenic fungi through mycobiome analysis in fungal infected FFPE tissues. However, we have found problems to be solved in further studies, such as increasing sequencing output, increasing fungal DNA concentration, excluding normal flora, and expanding fungal databases. If we develop a method to characterize pathogenic fungi in FFPE tissues in a follow-up study, we think it will help patients to use appropriate antifungal agents.

Data availability

Underlying data

Figshare: Mycobiome analysis in fungal Infected formalin-fixed and paraffin-embedded tissues for identification of pathogenic fungi: A pilot study. <https://doi.org/10.6084/m9.figshare.12616772.v1>²⁰.

This project contains the following underlying data:

- Unedited, original microscopy image files

Zenodo: byun1114/mycobiome_1: Mycobiome_anlaysis. <https://doi.org/10.5281/zenodo.3942711>¹⁹.

This project contains the following underlying data:

- FASTA.tar (sequencing data in FASTA format)

Extended data

Zenodo: byun1114/mycobiome_1: Mycobiome_anlaysis. <https://doi.org/10.5281/zenodo.3942711>¹⁹.

This project contains the following underlying data:

- SUPPLE_TABLE_1.csv (Supplementary Table 1 containing detailed BLAST results)

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

References

- Raja HA, Miller AN, Pearce CJ, *et al.*: **Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community.** *J Nat Prod.* 2017; **80**(3): 756–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aslam S, Tahir A, Aslam MF, *et al.*: **Recent advances in molecular techniques for the identification of phytopathogenic fungi - a mini review.** *J Plant Interact.* 2017; **12**(1): 493–504.
[Publisher Full Text](#)
- Cassagne C, Normand AC, L'Ollivier C, *et al.*: **Performance of MALDI-TOF MS platforms for fungal identification.** *Mycoses.* 2016; **59**(11): 678–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
- McNeil CJ, Luo RF, Vogel H, *et al.*: **Brain abscess caused by *Phaeoacremonium parasiticum* in an immunocompromised patient.** *J Clin Microbiol.* 2011; **49**(3): 1171–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Moncada PA, Budvytiene I, Ho DY, *et al.*: **Utility of DNA sequencing for direct identification of invasive fungi from fresh and formalin-fixed specimens.** *Am J Clin Pathol.* 2013; **140**(2): 203–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Benamu E, Yu AT, Xie L, *et al.*: ***Scedosporium apiospermum* infection of the urinary system with a review of treatment options and cases in the literature.** *Transpl Infect Dis.* 2018; **20**(1): e12804.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Trachuk P, Szymczak WA, Muscarella P, *et al.*: **A Case of Invasive Gastrointestinal Mycotypha Infection in a Patient with Neutropenia.** *Case Rep Infect Dis.* 2018; **2018**: 5864175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guarner J, Brandt ME: **Histopathologic diagnosis of fungal infections in the 21st century.** *Clin Microbiol Rev.* 2011; **24**(2): 247–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jung J, Park YS, Sung H, *et al.*: **Using immunohistochemistry to assess the accuracy of histomorphologic diagnosis of aspergillosis and mucormycosis.** *Clin Infect Dis.* 2015; **61**(11): 1664–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hong KH, Kim JW, Jang SJ, *et al.*: **Liver cirrhosis caused by *Exophiala dermatitidis*.** *J Med Microbiol.* 2009; **58**(Pt 5): 674–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Perfect JR: **The antifungal pipeline: a reality check.** *Nat Rev Drug Discov.* 2017; **16**(9): 603–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schoch CL, Seifert KA, Huhndorf S, *et al.*: **Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi.** *Proc Natl Acad Sci U S A.* 2012; **109**(16): 6241–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schoch CL, Robbertse B, Robert V, *et al.*: **Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi.** *Database (Oxford).* 2014; **2014**: bau061.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nilsson RH, Larsson KH, Taylor AFS, *et al.*: **The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications.** *Nucleic Acids Res.* 2019; **47**(D1): D259–D64.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shendure J, Balasubramanian S, Church GM, *et al.*: **DNA sequencing at 40: past, present and future.** *Nature.* 2017; **550**(7676): 345–53.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Branton D, Deamer DW, Marziali A, *et al.*: **The potential and challenges of nanopore sequencing.** *Nat Biotechnol.* 2008; **26**(10): 1146–53.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- White TJ, Bruns T, Lee S, *et al.*: **Amplification and direct sequencing of fungal ribosomal RNA genes For phylogenetics.** 1990.
[Reference Source](#)
- Camacho C, Coulouris G, Avagyan V, *et al.*: **BLAST+: architecture and applications.** *BMC Bioinformatics.* 2009; **10**: 421.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- byun1114: byun1114/mycobiome_1: Mycobiome_anlaysis (Version v1.0). Zenodo. 2020.
<http://www.doi.org/10.5281/zenodo.3942711>
- Byeon S, Lee T, Na HY, *et al.*: **Mycobiome analysis in fungal Infected formalin-fixed and paraffin-embedded tissues for identification of pathogenic fungi: A pilot study.** *figshare.* Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.12616772.v1>
- Santos ARO, Leon MP, Barros KO, *et al.*: ***Starmerella camargoi* f.a., sp. nov., *Starmerella ilheusensis* f.a., sp. nov., *Starmerella litoralis* f.a., sp. nov., *Starmerella opuntiae* f.a., sp. nov., *Starmerella roubikii* f.a., sp. nov. and *Starmerella vitae* f.a., sp. nov., isolated from flowers and bees, and transfer of related *Candida* species to the genus *Starmerella* as new combinations.** *Int J Syst Evol Microbiol.* 2018; **68**(4): 1333–43.
[PubMed Abstract](#) | [Publisher Full Text](#)
- De Filippis F, Laiola M, Blaiotta G, *et al.*: **Different Amplicon Targets for Sequencing-Based Studies of Fungal Diversity.** *Appl Environ Microbiol.* 2017; **83**(17): e00905–17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Eltzschig HK, Carmeliet P: **Hypoxia and inflammation.** *N Engl J Med.* 2011; **364**(7): 656–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cho M, Ahn S, Hong M, *et al.*: **Tissue recommendations for precision cancer therapy using next generation sequencing: a comprehensive single cancer center's experiences.** *Oncotarget.* 2017; **8**(26): 42478–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weirather JL, de Cesare M, Wang Y, *et al.*: **Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis.** *F1000Res.* 2017; **6**: 100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Spencer DH, Sehn JK, Abel HJ, *et al.*: **Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens.** *J Mol Diagn.* 2013; **15**(5): 623–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 02 December 2020

<https://doi.org/10.5256/f1000research.27729.r75308>

© 2020 Wang C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chengqiang Wang

Shandong Engineering Research Center of Plant-Microbial Restoration, Shandong Agricultural University, Tai'an, China

The authors have identified the mycobiome analysis to studying pathogenic fungi in fungal infected FFPE tissues by simple nanopore sequencing. It is a method to identify microbiological compositions in different samples, although the accuracy of this method is not sufficient for diagnosing pathogenic fungi. The experimental treatment and result analysis of this paper are permissible. Overall, the paper fits the scope of the journal. I just present some of my opinions for thinking.

1. I think that all ITS1 or ITS2 base information is more accurate for fungal identification, and so, a part of ITS1 or ITS2 sequences might be false positive.
2. The method used in this paper is suitable for genres identification, and so, species identification is not sufficient. For example, "two *Candida* species are similar in ITS sequences". I suggest using characteristic genes of species for species identification.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Analysis of microbial diversity. Molecular and genetic biology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 04 August 2020

<https://doi.org/10.5256/f1000research.27729.r67771>

© 2020 Faour W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wissam Faour

School of Medicine, Lebanese American University, Byblos, Lebanon

The authors suggested genomic tools to assess fungal infection by manipulating DNA extracted from paraffin-embedded tissue blocks. There is a growing importance of using FFPE tissues further in clinical assessment, therefore, the study continues to contribute to this field. The paper fits within the scope of the journal

1. The paper is well-written in good English.
2. The authors need to show the concentrations of the recovered DNA and how they quantify it.
3. How the authors controlled for damaged DNA in sequences negative for fungal DNA.
4. The study lack proper positive and negative control (can use FFPE known to be free of fungal infection).
5. Maybe showing a photo of the 1% non-denaturing agarose gel is useful to assess the quality of the isolated DNA.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Inflammation, Cell signaling, cell biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 05 Aug 2020

Sun-ju Byeon

Thank you for your kindly comments.

Q: The authors need to show the concentrations of the recovered DNA and how they quantify it.

A: We measured the concentration of the extracted DNA using Nanodrop. We described using Nanodrop in Page 8, but omitted the description in the method.

Q: How the authors controlled for damaged DNA in sequences negative for fungal DNA.

A: We have not been able to conduct further experiments on cases where fungal DNA was not found. The research fund we received for this study is not large, and the research fund has been exceeded to purchase MinION devices and reagents needed for sequencing. We discussed in a discussion how to increase the likelihood of detecting fungal DNA because it is not a condition to conduct additional experiments. At the end of this study, we plan to apply for research funds for follow-up studies that can complement the limitations in this study.

Q: The study lack proper positive and negative control (can use FFPE known to be free of fungal infection).

A: We evaluated the adequacy of the experiment using Candida (labeled #1-11) and Aspergillus spp. (labeled #1-12) cultured in the Department of Laboratory Medicine as a positive control. We uploaded the FASTA file for the this sample (<https://doi.org/10.6084/m9.figshare.12762431.v2>). Since ITS1~3 primer can amplify human DNA, a separate negative control was not set.

Q: Maybe showing a photo of the 1% non-denaturing agarose gel is useful to assess the quality of the isolated DNA.

A: We uploaded an agarose gel image of genomic DNA (<https://doi.org/10.6084/m9.figshare.12762422.v1>).

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research