Optimization and Application of Genotyping and Identification Techniques for *Pneumocystis jirovecii*

Lana Pasic
BSc (Immunology and Microbiology)
MScMed (Infection and Immunology)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Sydney Medical School
Faculty of Medicine
The University of Sydney
September 2021

Statement of Originality

This thesis is submitted to The University of Sydney in fulfillment of the requirements for the degree of Doctor of Philosophy.

I, Lana Pasic, certify that to the best of my knowledge the work presented in this thesis is original, except were acknowledged in the text.

I certify that all work described in this thesis was conducted by myself, except when otherwise indicated. It has not been submitted before for any degree or examination to any other institution.

Parts of this thesis have been published in the candidate's name, and authorship attributions for each manuscript are outlined on the page preceding it. In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Signed,

Lana Pasic

27 September 2021

Acknowledgments

As this journey comes to an end, I would firstly like to take this opportunity to thank the key people who have supported and helped me throughout this process, without each of you, this would have been an impossible task.

First and foremost, I would like to thank my primary supervisor, Professor Wieland Meyer, for allowing me into the fantastic fungal family, and making this feat of mine a reality. Thank you for your dedication, your passion has inspired me, and all the memories formed with the team, from Camperdown to Adelaide, Rio de Janeiro to Amsterdam - I have felt proud to represent the group and your mentorship. Thank you for your support and belief, as well as the endless hours and resources you have personally invested to truly kickstarting my scientific career.

My utmost gratitude also goes to my secondary supervisor, Dr Laszlo Irinyi. You have supported me since day 1, and I will always be thankful for your kindness, dropping whatever you were working on to come to help with whichever issue (minor or major) I had that day, advice, complete patience whenever teaching me and your friendship. Thank you for all the hard work you have put into mentoring me over the years.

From the Medical Mycology group, a big thank you to Alex Kan for all your help, often last minute and in times of crisis, for all your work helping the database come to life, patience, endless cakes and teaching me how to reset my password every 6 weeks. Without you, my sequences never stood a chance in organisation! Also, thank you to Krystyna Maszewska, a true key member of the group, for always assisting me in the lab, the delicious Polish Sliwka, for the many chats in the lab.

Thank you to my international collaborators, including Professor Sergio L. Vargas and Carolina Ponce from the University of Chile and Professor Maite Teresa Martin-Gomez and Dr Lidia Goterris from the Hospital Universitari Vall d'Hebron. Without your instant friendship,

kindness and support, our work together truly stands to make a significant impact in the community!

Thank you to “The Veterinarians” Professor Mark Krockenberger and Professor Richard Malik, even though I was not even in your school, you still supported, believed, and helped me whenever I reached out for help, no matter how big the favour. Your kindness and willingness to go the extra mile will not be forgotten.

A thank you to Professor Tania Sorrell, Associate Professor Sharon Chen, Dr Catriona Halliday, Associate Professor Sebastiaan Van Hal, Yiheng Hu, Menuk Jayawardena, Indira Basu, Wendy McKinney, Dr. Arthur Morris, John Rathjen, Dr. Eric Stone, Dr. Benjamin Schwessinger and Professor Rosely Zancope-Oliveira for the roles you played in helping this work come together, including your guidance, assistance, or sample donations.

A huge thank you to the kind staff and students from Level 3 and 4 from The Westmead Institute over the years, especially to Ali, Alma, Cristina, Darci, Diana, Dom, Erik, James, Helen and Moo Noi. The friendships we cultivated, many adventures together, Starbucks trips and long lunches helped make this a fun and wonderful ride.

Three students especially played a very influential and important role in my PhD, and by the end of this, I had left with three wonderful lifelong friends and unforgettable memories. Thank you, Dr. Laura Schmertmann and your sidekick Frida, Dr. Christiane Prosser, and (soon-to-be Dr.) Minh Thuy Vi Hoang. I am almost certain that if it were not for your endless kindness, friendship, never-ending chats between desks, which continued to endless online chatter, trips, and wine nights, I would not have come out the other end in one piece.

To my non-PhD family and friends, thank you for all the support over the years, the belief that I could do this and reminding me of life outside the PCR room. To my bestie Laura, thank you

for your logical and rational brain when I was panicking, always talking me off the ledge and showing me the meaning of true friendship.

Finally, a huge thank you the four people who formed my biggest support system. Thank you to my mum, dad, Sabrina (and Lady) for believing and pushing me for over a decade of studying, the sacrifices, love, and constant support in every aspect possible allowed me to follow this dream. Thank you to my Miki, my personal cheerleader, who packed up his entire life and travelled across the other side of the planet to hold my hand through the toughest troughs andputting your life on hold so I could complete the dream. Thank you, thank you, thank you!

The weight of the world is always easier when the burden is shared, and I could not have done it had I not shared it with you all. Thank you!

Abstract

Pneumocystis species are obligate fungi capable of causing opportunistic infection called Pneumocystis Pneumonia (PCP) in those who are immunocompromised, often resulting in serious disease and caring a high mortality rate. Due to host specificity and inability to conduct *in vitro* studies, molecular methods have stepped in to undertake the bulk of epidemiological studies.Despite this, many unknows regarding the fungus still exist, and there is little concordance between research groups. This thesis aimed to improve and standardize the identification and genotyping techniques currently used to enable the global research community to bridge gaps collaboratively. Chapter 1 outlines the current knowledge on *Pneumocystis* epidemiology, the pitfalls facing researchers and how they have been addressed. As a major influencing factor in determiningsequencing success is the quality of the *Pneumocystis* DNA obtained, Chapter 2 investigated and compared several commercial DNA extraction kits and methods to determine an effective protocol to be used on lower respiratory samples. The gold standard for *P. jirovecii* currently remains to be multilocus sequence typing (MLST), therefore Chapter 3, optimized primers and PCR protocols of the current MLST scheme used by The International Society for Human & Animal Mycology (ISHAM). It identified pitfalls within the scheme and addressed them by reviewing all genotyping loci and schemes used till now to determine the most effective scheme to be used in the future. This resulted in the proposal of a global consensus MLST scheme for *P. jirovecii*, which was then successfully applied to a clinical cohort in Chapter 4. Chapter 5 examines the epidemiology of globally collected clinical *P. jirovecii* positive samples using the consensus scheme, uncovering unique alleles that formed 49 sequence types, and demonstrated the importance of a collaborative database. The importance of the genetic locus DHPS to monitor global resistance is discussed in Chapter 6. Finally, long-read next-generation sequencing using the Oxford Nanopore MinION sequencer is applied to detect *P. jirovecii* and *P. canis* in clinical samples to determine its potential for early diagnostics of PCP in Chapter 7. Chapter 8 is summarizing the finding made in this PhD research and provides future directions in this research field.

List of Abbreviations

Abbreviation	Meaning
26S rRNA	Large subunit of the rrna gene
AIDS	Acquired immunodeficiency syndrome
AT	Allele type
β-TUB	B-tubulin
BAL	Bronchoalveolar lavage
CKCS	Cavalier king charles spaniels
Ct	Cycle threshold
CT	Computed tomography
CW	Calcofluor white
CYB	Cytochrome b
DNA	Deoxyribonucleic acid
DHFR	Dihydrofolate reductase
DHPPP	6-hydroxymethyl-7,8-dihydropterin pyrophosphate
DHPS	Dihydropteroate synthase
dNTP	Dinucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
FAS	Folic acid synthesis
GSM	Grocott-Gomori methenamine-silver
GTR	General time reversible
HAART	Highly active anti-retrovirus therapy
HIV	Human immunodeficiency virus
IA	Index of association
ICU	Intensive care unit
IFA	Immunofluorescence antibody
IS	Induced sputum
ITS	Internal transcribed spacer
KEX1	Serine endopeptidase
ML	Maximum likelihood
MLEE	Multilocus enzyme electrophoresis
MLST	Multi-locus sequence typing
mm^3	Cubic millimetre
MRI	Magnetic resonance imaging
MSSU	Mitochondrial small subunit
NGS	Next generation sequencing
NPA	Nasopharyngeal aspirate/washes
NPS	Nasopharyngeal swabs
NSW	New south wales
OPW	Oropharyngeal wash
OTU	Operational taxonomic unit
PABA	Para-aminobenzoic acid
PBS	Phosphate-buffered saline
PCP	Pneumocystis pneumonia
PCR	Polymerase chain reaction

Qo	Quinol oxidation
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
SCID	Severe combined immunodeficiency
SD	Standard deviation
SOD	Superoxide dismutase
SSCP	Single strand conformation polymorphism
SSR	Simple sequence repeats
ST	Sequence type
TRR1	Thioredoxin reductase
TS	Thymidylate synthase
TSOH	Type-specific oligonucleotide hybridization
UCS	Upstream conserved sequence
WGS	Whole genome sequencing

Table of Contents

Chapter 1

Introduction

1.1 History

Pneumocystis was first identified by C. Chagas in 1909 [1]. Chagas was working on the parasitic *Trypanosoma* species in guinea pig models and assumed it was another trypanosome life form that was disoved [9]. Similarly, a year later A. Carinii noted the same organism in lungs of ratswhich he also thought to be *Trypanosoma* [2]. It was only in 1912 while working at the Institute Pasteur, that M. Delanoe and P. Delanoe realized it was an entirely new organism, noting that and subsequently naming it *Pneumocystis carinii* [3]. The name was derived from the pulmonarypathogenesis of the organism, and also in honour of A. Carinii [4]. In the 1940's, physicians around Europe had begun describing pneumonia of unknown origin occurring in malnourished infants and elderly populations in nursing homes. By the 1960's pneumonia caused by the organism *Pneumocystis*, *Pneumocystis* pneumonia (PCP), was recognized as an opportunistic disease. PCP was reported largely from affect patients with severely compromised immune systems, such asthose suffering from acute leukaemia or those with T-lymphocyte impairment [5]. In 1976, J. Frankel noticed phenotypic differences between the *Pneumocystis* infecting mammals and those in humans. Frankel renamed the organism, which infected solely humans, to *Pneumocystis jirovecii*. The name was selected to honour pathologist O. Jirovec who first reported it from humans in 1951 [6, 9].

It was not until the rise of the AIDS epidemic, that *Pneumocystis jirovecii* really received large-scale recognition as an opportunistic pathogen [7]. Before 1981, PCP was a relatively rare disease, but in 1981, clusters of adult males were appearing in hospitals around the United States, showing PCP symptoms. Most of the patients were males who had sexual intercourse with males, orinjecting drug users, but had no previous medical predispositions or any known causes of impaired immunity [8]. This initial PCP outbreak helped with the later identification of HIV and subsequently AIDS. PCP was the most common AIDS defining illness in the developed world. To

this day, PCP remains a leading opportunistic infection, with high levels of morbidity and mortality around the globe, for both HIV and non-HIV patients with severe immunodeficiency [9, 10].

1.2 Taxonomy

Pneumocystis organisms are unique and complex, with an interesting taxonomic history. Originally, the class of organisms was considered to be a protozoan. This was hypothesized on the basis of several similar morphological characteristics, and their susceptibility to pentamidine, an antiprotozoal drug [9]. It was not until 1988 that *Pneumocystis* was shown to belong to the fungal kingdom. This discovery came from Edman and Stringer, who sequenced the ribosomal RNA and found it was not related to protozaons, but instead to fungi [11]. Ever since, all genomic analysis has agreed with this finding. To date, *Pneumocystis* is classified, as below, in the phylum *Ascomycota*, but in their own unique class, order and family [2]:

Kingdom: *Fungi*

Phylum *Ascomycota*

Class: *Pneumocystidomycetes*

Order: *Pneumocystidales*

Family: *Pneumocystidaceae*

When it was first identified that the species varied, *Pneumocystis* species were named using *forma specialis* (f. sp.) depending on what mammal the organism was infected. The human type was called *Pneumocystis carinii* f. [or f. sp.] *hominis*, the mice type called *Pneumocystis carinii* f. [or f. sp.] *murina*, and the original type which infects rats called *Pneumocystis carinii* f. [or f. sp.] *carinii*. This naming system is still used by some to this day, but is now less common [12, 13].

Although in the late 1970's the name *Pneumocystis jirovecii* was coined for the specific species, which infected humans, it was not until the early 21st century that Stringer *et al.* proposed the use of this name in the scientific literature, and the International Code of Botanical Nomenclature accepted it [6, 14].

The renaming of *P. carinii* to *P. jirovecii* is still subject to controversy and has not been unanimously accepted [15]. Further debate arose when there was a proposed change of PCP to PJP to better suit the new name change of the organism causing pneumonia in humans. Due to the widespread use of PCP in both the scientific and medical world, it was decided that PCP would remain, but now stand for **P**neumo**C**ystis (*jirovecii*) **P**neumonia, rather than *Pneumocystis carinii* Pneumonia [9, 15, 16].

Pneumocystis is said to be monoxenous, meaning different mammals have been shown to have their own specific species of *Pneumocystis*, some theories even suggest there may be a species of *Pneumocystis* for each mammal [17]. Each species is extremely host specific, with differing genetics that have co-evolved with the host species and are unable to cross-infect other species [9, 18]. Despite this, to date, only 5 species have been officially named from the f. sp. form, including *P. carinii* as originally named from rats (Rattus norvegicus), *P. murina* from mice (*Mus musculus*), *P. wakefieldiae* also from rats, and *P. oryctolagi* from rabbits (*Oryctolagus cuniculus*) [12].

Molecular analysis showed that there was a difference of 5% between human *P. jirovecii* and rat *P. carini* when looking at the 18S rRNA sequences (Figure 1.1) [9, 19]. Another study found that *Pneumocystis* from pigs differed greater than 15% when examining the *mt26S* locus [20]. This is an unusually high level of divergence between species when compared to other fungal genera that also evolved throughout the same time-period – estimated to be tens of millions of years ago [21]. When comparing DNA sequences, genetic differences are evident within all organisms of the genus *Pneumocystis*, with studies showing up to 7% to 35% difference in homologues genes, and a 30% divergence occurring in sequences between specific genes.

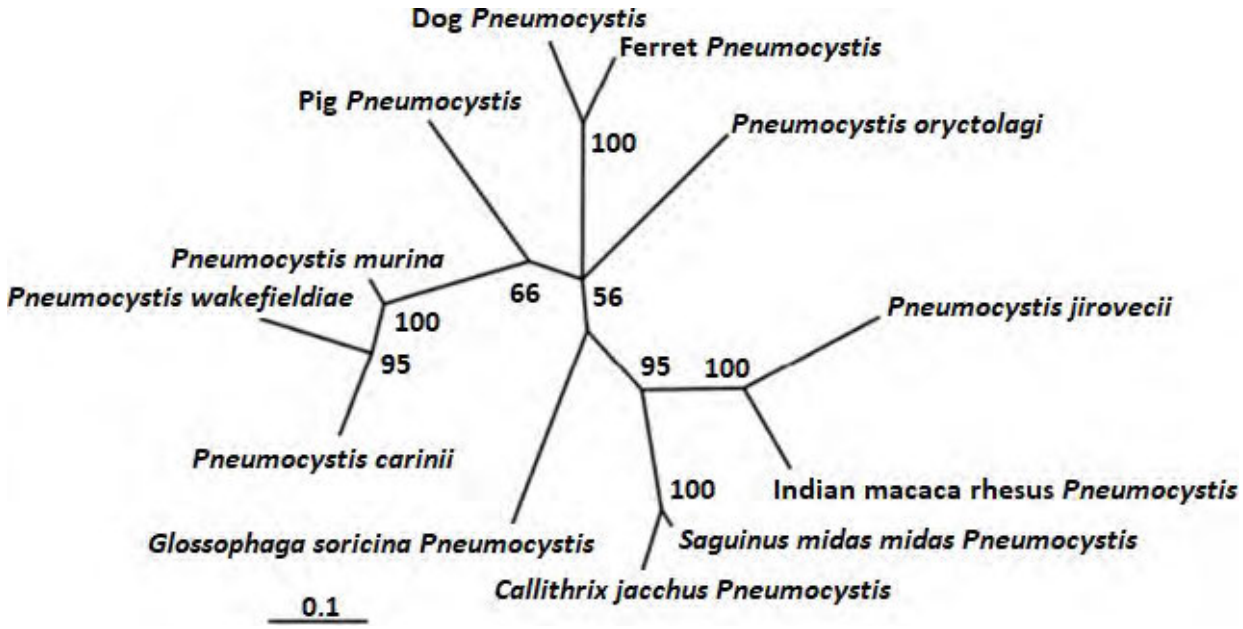


Figure 1.1. *Pneumocystis* species phylogenetic tree. Phylogenetic tree of various *Pneumocystis* species, adapted from Magali Chabé *et al.*, 2010, displays genetic distinction of species, show by analysis of 18S rRNA sequences [22]. Analysis shows 5% distinction between species for the 18S rRNA sequence, while some coding regions have displayed even higher sequence differentiation.

A study by Cisse *et al.*, 2018, shows that the molecular clock estimates indicate that *Pneumocystis* species diverged before their host (Figure 1.2) [24]. Although this strongly supports the hypothesis that *Pneumocystis* is monoxenous, some studies have now also started questioning this hypothesis [23]. This idea was first introduced by Guillot *et al.*, 2004, when an identical strain of *Pneumocystis* successfully infected two separate species of macaques, rhesus and long tailed, andhas since been reinforced by a handful of examples, such as two rat species which are separated by7 million years of evolution harbored the same *Pneumocystis* genetic sequence [25]. This school of thoughts suggests that the time of evolutionary host divergence guides the specificity, rather than only speciation. Further *Pneumocystis* sequence data are needed, from a range of host mammals, to help further to understand the evolutionary history and speciation of this fungi. There is currently no consensus explanation for these specific requirements, which is largely unseen in the fungal kingdom. [7, 14].

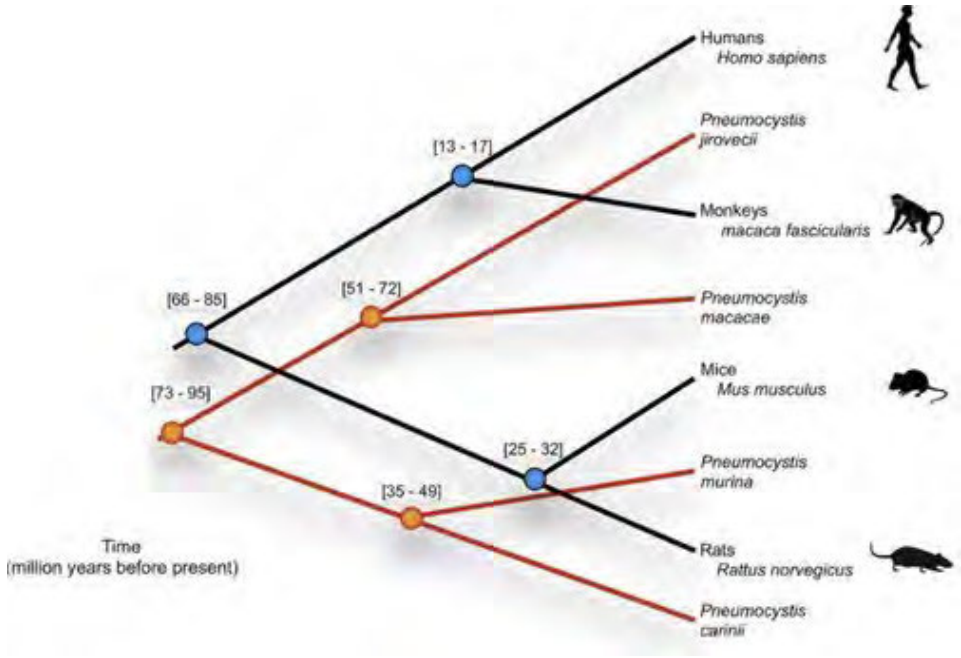


Figure 1.2. Co-evolution of *Pneumocystis* species. Evolutionary visualisation by Cisse *et. al.*, 2018, showing the molecular clock estimates which indicate that *Pneumocystis* species diverged before their respective hosts.

1. 3 Problems with cell cultivation

To date, *Pneumocystis* has not been able to be cultured or propagated in environments outside of mammalian lungs, which to many is seen as a huge drawback for researchers. Specifically, *Pneumocystis jirovecii* has not been cultured, nor can it be grown *in vitro* in any other animal [26]. *P. jirovecii* has only been successful proliferated transiently in short term, which was in within lung epithelial cells in tissue culture medium. Although long term continued passage and contiunal growth has been attempted by replicating an alveoli-like environment, but none have yet been successful [26, 27]. In 2014, Verena Schildgen *et al.* 2014, stated that *P. jirovecii* can be successfully cultured and propagated from *P. jirovecii*-positive bronchoalveolar lavage (BAL) patient samples by using a permanent three-dimensional air-liquid interface culture system formed by CuFi-8 cells [28]. This excited many *Pneumocystis* study groups, but unfortunately this was unable to be replicated by any other research groups so far [29]. Due to these restrictions and the inability to propagate, studies have been based mainly on light and electron microscope analysis, and molecular analysis. [9, 11, 30].

1. 4 Life Cycle

Pneumocystis has a biphasic life cycle, consisting of two morphological forms: the trophic and cyst form (or ascus). The trophic form is smaller, measures 1 to 4 μm in diameter, while the most mature cyst formis larger, usually 8 to 10 μm in diameter (Figure 1.3) [7]. The trophic form has a simple plasma membrane and often filopodial projections, while a rigid rich b-glucan wall characterizes the cyst form [7]. During infection, the trophic form is 10 times more common in the lung than the cystform. The cyst form is said to play a major role in *Pneumocystis* propagation.

Motile haploid trophic forms of *Pneumocystis* have been hypothesized to bind tightly to lung alveolar epithelial cells [26]. If the attachment is prevented, *Pneumocystis* cannot infect, propagate, nor proliferate. The attachment is followed by a sexual conjugation-mating phase, resulting in a cyst. The cyst will undergo both meiosis and mitosis, until finally becoming a mature cyst [13, 30].

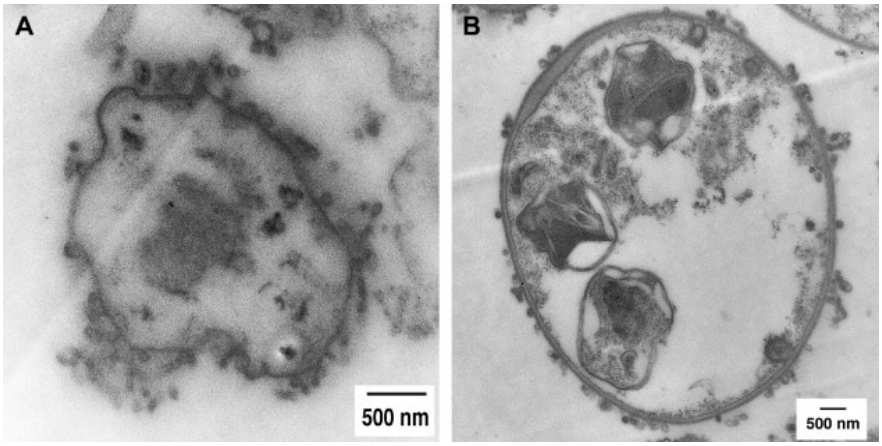


Figure 1.3. Life cycle forms of *Pneumocystis.* Electron microscope images of *Pneumocystis* trophic forms (A) and cyst form (B). The trophic forms are smaller, with filopodial projections. The cyst form is larger, rounder and is characterized by a b-glucan wall. It may have up to eight intra-cystic nuclei. Image from Krajicek *et al.*, 2009 [7].

There have been three intermediate cyst stages visualized, early, intermediate, and late, before they become mature cysts. The intermediate cyst stages start with complements of two nuclei, doubling to four, and then eight, respectively. Ultimately the mature cyst contains eight intra-cystic nuclei, also referred to us ascospores [31, 106]. When the cyst ruptures, the haploid trophic form is again produced (Figure 1.4) [32].

The trophic form may then go through vegetative growth or sexual conjugation again. It has also been proposed that the trophic form can undergo asexual reproduction, via haploid mitosis and binary fission [7, 32]. However, asexual reproduction still remains a matter of debate and study (Figure 1.5) [26]. This life cycle model is compatible with models of ascomycetous fungi [9]. If one day, *Pneumocystis* will be able to be cultured *in vitro*, many more insights and understandings ofthe life cycle will be available, but until then, the most reliable data will be from electron microscopy.

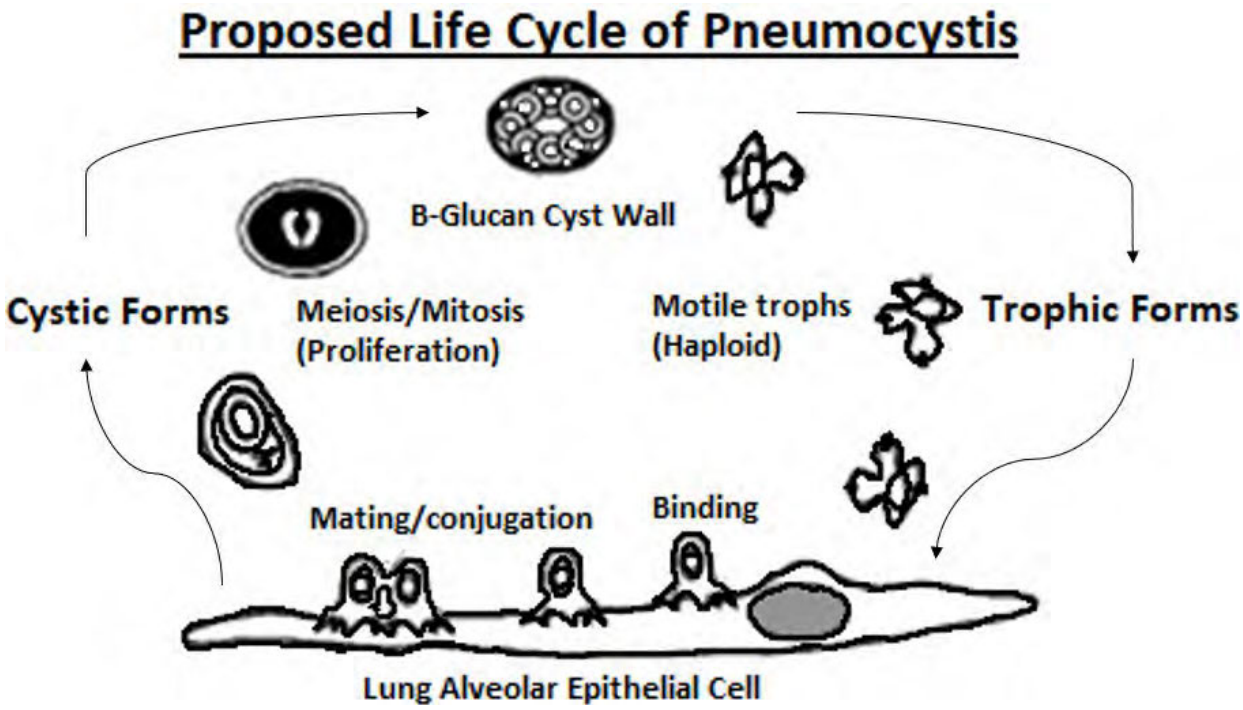


Figure 1.4. Proposed Life Cycle of *Pneumocystis*. The trophic form attaches to the host cell, where sexual conjugation takes place, forming the cyst form of the organism. The cyst matures, and new trophic forms are released. Diagram altered from Krajicek *et al.*, 2009 [7].

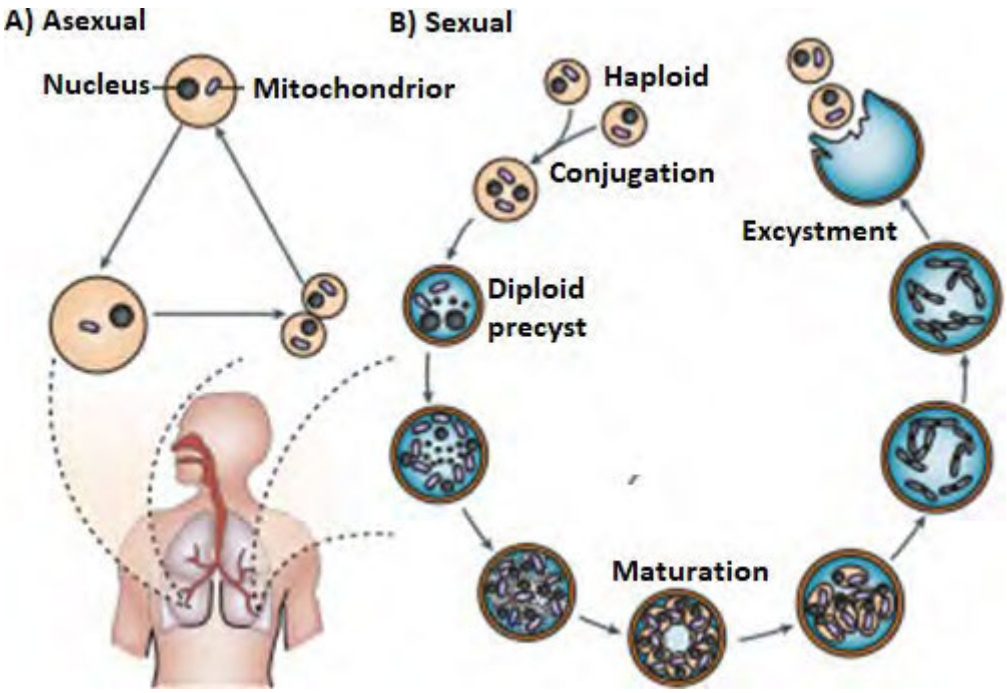


Figure 1.5. Sexual VS Asexual Life Cycle of *Pneumocystis*. A) Describes the asexual reproduction of *Pneumocystis*, which involves conjugation via binary fission. B) Shows the sexual reproduction of the cyst development stage. The cyst develops into 2, then 4, and then 8 inter-nuclei cyst before maturing and rupturing for the release of the trophic forms. Diagram originally from Thomas and Limper, 2007 [26].

1.5 Cell Interactions and Host Response

It is currently understood that *Pneumocystis* can survive and replicate almost exclusively in the alveoli of the lung, interestingly, disseminated PCP has been reported once in 1998, in extremely rare cases of serious infection in the context of severe immunosuppression [13, 26, 102]. The trophic form of the organism will attach and bind to alveolar epithelial cells by using fibronecin and vitonectin, which are presentin the lung fluid. To ensure attachment, *Pneumocystis* uses receptors and integrins to coat its surface with the proteins [33]. Adherence to the cell will trigger a cyclin-dependent kinase regulatory pathway, which will actively inhibit further growth of host epithelial cells [34]. A study by C. Thomas and A. Limper in 2007 showed that invasion of the actual host cell itself is uncommon [26]. *Pneumocystis* is believed to maintain an extracellular existence within alveoli, leaving the epithelial cell structure and barrier function undisrupted. Although, once attachment is secured, *Pneumocystis* proliferation is strongly promoted. The attachment to epithelial cells also initiates a

cascade of cellular responses, via selective kinase signaling pathways, in both host cells, and the organism [9, 13, 26].

Macrophages are the primary resident phagocytes of the lung, which mediate pathogen clearance [9, 12]. Attachment of *Pneumocystis* will promote the production of chemokines and inflammatory cytokines by the macrophages, also initiating inflammation. Once macrophages internalize the organism, it is incorporated into phagolysosomes, and in turn degraded and cleared [9]. Studies showed that rodent models, which had a marked depletion in macrophages, showed severe impairment in the clearance of *Pneumocystis*, compared to those of normal macrophage levels [35]. This offers insights as to why the pathogenesis of PCP is heightened in AIDS patients [111, 112, 113, 114].

A host immune response against PCP also includes the involvement of neutrophils and T-lymphocytes. CD4+ T cells have an important role in the control of the infection, especially in the recruitment of extra immune effector cells such as monocytes and macrophages [9, 36]. Additionally, IgG can be recruited for the opsonization of the pathogen [26]. Studies showed that mice with Severe Combined Immunodeficiency (SCID) could only control and clear a *Pneumocystis* infection once CD4+ T cells were re-introduced in the mice [37]. Although an effective host immune response is necessary to fight and ultimately clear a PCP infection, inflammation during the infection is also the cause of pulmonary damage. In severe infection, inflammation is intensified with the excess involvement of immune cells, such as neutrophils and CD8+ T cells. In SCID mice, the reintroduction of CD8+ T cells did not offer protection against infection, but instead caused additional alveolar impairment [37]. Furthermore, the excess recruitment of neutrophils, via CXCL2 and IL-8, correlated highly with lung damage. Excess neutrophils and CD8+ T infiltration will create alveolar damage, and cause impaired gas exchange, which can lead to respiratory failure, and death [36, 38]. Respiratory impairment and the mortality related to PCP are often directly related to the amount of inflammatory damage in the lungs, rather than organism burden of *Pneumocystis* [39].

1.6 Clinical Symptoms and Presentation of PCP

Diagnosis of PCP from a physical examination by primary care physicians can be quite challenging, as PCP presents with non-specific signs and symptoms, often presenting with standardized pneumonia symptoms [9]. Symptoms usually include a fever, a non-productive cough or a cough producing clear sputum, arterial partial pressure of oxygen below 65 mmHg, malaise and dyspnea [40, 83]. Furthermore, regardless of the level of infection of hypoxemia, pulmonary auscultation is often normal, with only discrete crackles present sometimes [41].

Although PCP is known as an AIDS defining illness, it also occurs in non-HIV patients, which are severely immunocompromised, such as organ transplant recipients, cancer patients and autoimmune patients. PCP presents differently in HIV and non-HIV patients [41]. In HIV patients, PCP will develop in more than 60% of patients during the disease course, but often only present when their CD4+ T cells are below 200 cells per cubic millimeter (mm^3) [42, 83]. Symptoms in HIV patients will often progress over a few weeks, usually 25 to 28 days, developing a subacute onset of disease. HIV patients also show with higher arterial oxygen tension and lower alveolar- arterial oxygen gradient [30, 43, 44]. Conversely non-HIV patients will produce a more acute onset of disease, typically within 5 to 6 days, with much more aggressive respiratory failure and more often needing mechanical ventilation, than HIV patients [43, 44].

A chest radiograph presents with similar features between HIV and non-HIV patients (Figure 1.6). It usually shows bilateral, diffuse, reticular, or granular opacities, which become more homogenous and diffuse as the infection progresses and becomes more serious. There are also many other less common and non-specific patterns, which have been previously featured on the radiograph [45]. In the last decade, the majority of patients will have evidence upon imaging [103]. A CT scan typically is more sensitive and will yield more characteristic featuresof a PCP infection, even when radiograph findings are normal. Some of these features include ground-glass opacities with patch distribution in perihilar regions of the lungs, along with thickened

septal lines and areas of consolidation [9, 46]. Furthermore, cysts can be observed in up to 40% of patients [9].

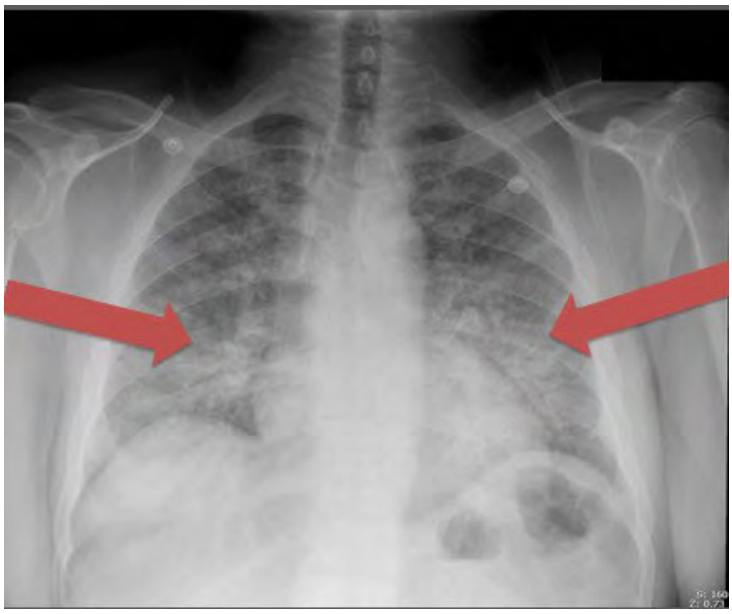


Figure 1.6. Chest Radiograph of PCP Infection. PCP shows bilateral perihilar interstitial infiltrates (red arrows) that become more extensive and may progress to consolidations over several days. Image from Medscape, 2018 [47].

It has been found that up to 20% of adults have *P. jirovecii* DNA detected in their lungs but show no signs of clinical disease. When this occurs, individuals have been dubbed Pneumocystis colonization or carriers [9]. Asymptomatic carriage is a very important factor in the transmission of *Pneumocystis* [48]. It has been postulated that the primary infection is actually asymptomatic, but further analysis needs to be done to assess if there are actually subtle clinical manifestations occurring or not [49].

1.7 Diagnosis

Due to the ambiguous nature of clinical and radiological presentations of PCP, diagnosis strong depends on the visualization of the organism and further analyses. Respiratory specimens, such as

bronchoalveolar lavage (BAL) fluid, induced sputum (IS) or oropharyngeal (OPW) wash samples, are essential for the diagnosis of PCP [9].

Bronchoscopy with immunofluorescence antibody (IFA) stains of BAL is the preferred diagnostic procedure, with a high sensitivity between 89% and 98% [9]. Monoclonal anti-*Pneumocystis* antibodies used for staining stain both trophic forms and cysts, and in turn are much more sensitive than general stains. This technique remains the gold standard for PCP identification [30]. IS and OPW can also be used, and are less invasive to obtain, however, the sensitivity is markedly lower [50]. There are also other stains, which can visualize and help confirm *Pneumocystis* within a host. Trophic forms found in respiratory samples can be analysed using Papanicolaou, Gram-Weigert or modified Wright Giemsa stains; while Grocott-Gomori methenamine-silver (GMS), cresyl echt violet toluidine blue O stain, and calcofluor white (CW) canbe used to identify cysts (Figure 1.7) [9, 13, 51]. Although cheaper, quicker and more specific than IFA stains, many of the more conventional stains require very experienced staff to perform thestain, and also analyse the results [52]. A study, which reviewed these specific stains, found that Wright Giemsa stains, CW, and GMS were the most suitable for use in a clinical laboratory. Another study recommended the use of the more specific stains to act as a secondary confirmation stain, when IFA is used as the primary staining method, to heighten both specificity and sensitivity to maximum levels [51].

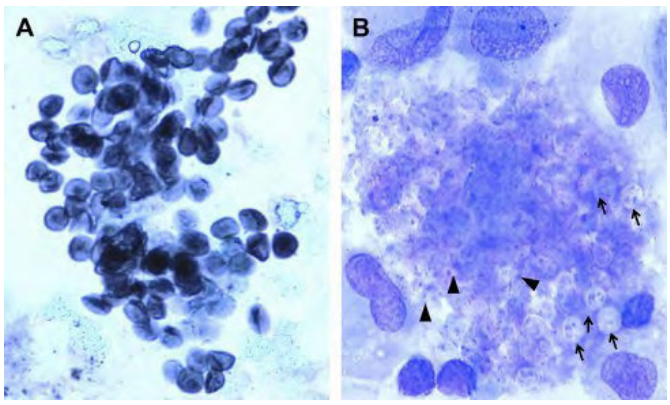


Figure 1.7. GMS and Giemsa Stains of *Pneumocystis*. *Pneumocystis* in BAL specimen stained with GMS (A), has stained the cyst walls silver, and is seen as oval bodies. Giemsa stain (B) shows both the trophic and cyst forms. Intra- cyst nuclei are seen in cysts (shown with arrows), and trophic form is identified by pale blue-grey cytoplasm and dot- like nuclei (shown with arrow heads). Original amplification X 500. Image taken from Catherinot *et al.*, 2010 [9].

For a microbiological diagnosis to be successful, it relies on several factors, such as (1) organism burden, (2) quality of the extracted sample, and (3) the competence of the staff performing the test [51]. This can be very problematic, especially in non-HIV patients, who have significantly lower levels of *Pneumocystis* than in AIDS patients. These patients will have a higher number of host leukocytes, and lower levels of the actual pathogen during an episode of PCP, making visualization of the organism in a respiratory sample extremely difficult, if not impossible [13, 26]. This complication has prompted the development of more precise molecular detection methods.

PCR detection methods have been developed, in attempts to circumvent low sensitivity and specificity in poor quality respiratory samples. Both conventional and quantitative PCR have been used, using various: target genes, PCR platforms, protocols, and assays [53]. PCR has higher specificity (86% - 100% from BAL samples) than microscopic detection methods (70 – 92% from BAL samples), but still has problems regarding low positive predictive values, and false positives of PCP diagnosis in patients with asymptomatic carriage [54-56]. Nested PCR increases the specificity in comparison to conventional PCR to greater than 80%. Quantitative real-time PCR shows even more promising results for use in clinical settings. It has a specificity of 96% and can be used for accurate quantification of organism burden, therefore also allowing the identification of potential *Pneumocystis* carriers, and distinguishing between colonization and clinical disease [55]. Studies also show that reverse transcriptase PCR assays have a heightened specificity at 86% and a sensitivity of 100%. Although less stable to use than DNA, RT-PCR uses messenger RNA and is able to give information on surrogate viability too [57]. At this point in time, molecular methods arebeing used much more often as clinical research tools, than solely for diagnostics.

Previously, in an extremely rare case, where all detection methods have yielded unsatisfactory results, an open lung biopsy has been necessary for diagnosis [58].

1. 8 Treatment and Prevention

The most recommended, effective, and wisely used agent in treating mild to severe PCP is trimethoprim-sulfamethoxazole (TMP-SMX), also called co-trimoxazole, in both HIV and non-HIV patients. It can be taken either intravenously or orally but can have adverse effects that usually begin in the second week of treatment. It is often prescribed for 14 to 21 days, depending on the patient and severity of the disease [59]. Alternative drugs, along with dosage and side effects are summarized in Table 1.1.

Neither the exact route of transmission nor the ecological niche of *Pneumocystis* is known, making prevention strategies for health care workers more difficult. Various chemoprophylaxis regimes have been suggested for both HIV and non-HIV individuals, depending on specific needs to protect the patient's immune system and their vulnerability [9]. TMP-SMX is also the first line of prophylaxis in both HIV and non-HIV patients. One double strength tablet daily is the favoured regimen, and has yielded very high and effective protection rates, often reaching 90% to 100% [60]. However, TMP-SMX does have side effects, which often make tolerating the drug difficult. The benefit of taking chemoprophylaxis should be balanced out with the risk of acquiring a severe side effect, and highly depend on the likelihood of the individual acquiring PCP [61]. Table 1.1 summarizes suggested TMP-SMX prophylactic regimes for certain patient groups.

Additionally, since nosocomial acquisition of disease has been previously noted in outbreaks, prevention of air transmission has been suggested to avoid secondary cases. Transmission between PCP patients can be stopped via respiratory isolation and keeping patients in single rooms [62].

When looking at the relationship between co-trimazole usage and DHPS mutations, since the 1990's there has been a growing increase of mutations in *P. jirovecii*, which may indicate resistance towards co-trimaxazole [109, 110]. In the more recent decade there has been a change in resistance patterns, discussed further in Chapter 6.

Co- trimoxazole targets and inhibits dihydrofolate reductase and dihydropteroate synthase (*DHPS*).Using sulfa and sulfone drugs extensively to prevent and treat PCP in the last two decades has resulted in an increase of mutations in the gene coding for *DHPS noted* [7, 63].

Resistance to the drug has also been linked to recurrence of PCP infection, making treatment options harder to manage for these patient groups. Screening of the *DHPS* gene for resistance has slowly been implemented in hospitals of endemic areas, to help practitioners combat the growing resistance [64, 65].

Table 1.1. Treatment and Prophylaxis Regimes for PCP. Table compiled from Catherinot *et al.*, 2010, Bucher *et al.*, 1997, and Stein *et al.*, 1991 [9, 60, 61].

	Drug of choice	Dosage for treatment	Dosage for prophylaxis	Adverse side events
First choice	Trimethoprim + Sulfamethoxazole	15–20 mg/kg 75–100 mg/kg Intravenous or orally, divided into 3–4 doses daily	80–160 mg daily with 400–800 mg daily orally or intravenous, OR 160 mg 3 times a week 800 mg 3 times a week orally or intravenous	Skin reactions, Hepatitis, Pancreatitis, Gastrointestinal disturbance, Renal insufficiency, Hyperkalemia
Alternative for a mild to moderate PCP infection or risk	Dapsone with Trimethoprime	100 mg/d orally 5 mg/kg 3 times daily	-	Methemoglobinemia, Skin rash, Fever, Gastrointestinal disturbance
	Atovaquone	750 mg 2–3 times daily orally	750 mg 2 times daily orally	Skin rash, Fever, Gastrointestinal disturbance, Hepatitis
	Clindamycin with Primaquine	600 mg 4 times daily intravenous or 350–400 mg 4 times daily orally 30 mg daily orally	-	Skin rash, Fever, Neutropenia, Gastrointestinal disturbance, Methemoglobinemia
	Dapsone	-	100 mg daily orally	Contraindication in cases of G6PD deficiency Possible cross-reaction with sulfa allergy
	Dapsone + Pyrimethamine	-	50 mg daily orally 50 mg per week + leucovorin 25 mg per week	Contraindication in cases of G6PD deficiency Possible cross-reaction with sulfa allergy
Alternative for a moderate to severe PCP infection or risk	Pentamidine	4 mg/kg daily intravenous	300 mg monthly	Hypotension, Cardiac arrhythmias, Hyperkalemia, Hypomagnesemia, Hypocalcemia, Renal insufficiency, Pancreatitis, Hypoglycaemia, Diabetes mellitus, Neutropenia, Hepatitis

1.9 Prognosis

PCP is a severe infection, with high morbidity and mortality [115, 116]. Despite treatment, PCP can be life threatening. Patients with HIV generally display better outcomes and chances of survival, than non- HIV patients, that have other underlining immunosuppressive conditions [43]. In a recent review survival in HIV patients was shown to be 86% to 92%, with an overall hospital mortality of 11.6%, and 29% mortality rate of HIV patients in ICU [66]. Non-HIV patients have a lower survival rate, between 50% and 80% [67]. Due to high mortality rates, prevention should be prioritized in high-risk individuals.

1.10 Transmission

There have been several methods of transmission of *P. jirovecii* among humans, including environmental exposure, air-borne acquisition from infected humans, and also reactivation of latent infections [26]. Since the environmental niche of *Pneumocystis* is still unknown, environmental exposure is hard to define [68].

Due to exposure to *Pneumocystis* early in life, for many years it was strongly believed that PCP infections were due to reactivation of latent infections. In more recent studies, this theory is proving to be less and less substantial [68]. A study looked at the analysis of PCP outbreaks inurban settings and discovered that current PCP infections in patients were related to their place of residence, opposed to their place of birth [69]. This theory was also argued against using animal model experiments. Mice infected with *Pneumocystis* cleared the infection within 3 weeks, and theydid not re-develop PCP when their CD4+ T cells were depleted [70].

Furthermore, in humans, *Pneumocystis* was completely eliminated after infection, and persistence of latent organisms was very limited [71]. This has promoted the theory of airborne

transmission, which has been favoured in both animal and human studies since. Rats infected with *Pneumocystis* were able to transmit the infection to germ-free rats they were either co-housed with, or even to rats housed in adjacent cages [72]. Additionally, molecular analysis of PCP outbreaks among renal transplant patients discovered the infections are mostly nosocomial related infections, and to distinctively differ from other cases in different hospitals, or in the same hospital but in a different time frame [73]. In another study, air from the room of PCP patients yielded positive results when tested for *Pneumocystis* DNA. *Pneumocystis* DNA was also detected in the respiratorytract of healthy individuals after exposure to PCP patients, and in the air of PCP patient's rooms[74, 75]. It was shown that a close-contact period as short as 1 day was sufficient enough totransmit the infection [62].

Additionally, the colonization of healthy individuals may play a huge role in the transmission cycle by becoming a reservoir for *Pneumocystis* transmission. Reservoirs of the fungus can transfer *Pneumocystis* from carrier to carrier, without knowledge or any clinical symptoms. This keeps the organism active and circulating in the environment until it reaches an immunocompromised host and PCP is developed [48, 62].

Lastly, there has also been limited evidence of vertical transmission from mother to child, with possible transmission occurring via the transplacental route [107, 108].

1.11 Epidemiology and Prevalence - Risk Factors

PCP is one of the most prevalent opportunistic infections in AIDS and severely immunocompromised patients. It has a worldwide distribution and is detected in the environment around apple orchards and ponds, but no specific environmental niche has been confirmed [68, 76]. Exposure to *Pneumocystis* is common early in life, and studies have shown that PCP will present itself as a self-limiting infection in most children, usually before the age of 9 months [49, 76]. A high seroprevalence of anti-*Pneumocystis* anti-bodies seen in healthy children worldwide reinforcesthis notion [77]. Although the prevalence of these antibodies varies depending on the geographic

region, a study found that anti-*Pneumocystis* antibodies could reach a prevalence of 85%, in healthy children, by the age of 13 [76]. Along with the theory of airborne *Pneumocystis* transmission, these healthy carriers, with short-lived asymptomatic *Pneumocystis* colonization, are now regarded to be potential reservoirs for risk groups of developing clinical PCP.

Some individuals have a greater risk of acquiring PCP colonization, which may develop into clinical infection, given the right circumstances. *Pneumocystis* colonization can range to up to 68% in HIV patients, on highly active anti-retrovirus therapy (HAART) regimes and receiving anti-*Pneumocystis* prophylaxis [78].

Other medical conditions or factors that promote a higher risk of *Pneumocystis* colonization include chronic lung disease, diabetes, myeloma, leukaemia, solid malignancies, sarcoidosis, whole organ or bone marrow transplant recipients, primary immunodeficiencies or autoimmune diseases, collagen-vascular disorders, receiving corticosteroid treatment, receiving immunosuppressive medication, pregnancy, and even smoking [9, 26, 77]. These risk groups are more susceptible to colonization but may only act as a reservoir for the organism, without developing PCP.

Risk groups with a heightened chance of developing PCP, are urged to undergo prophylaxis regimes. PCP will manifest in HIV patients predominantly when their CD4+ T cell count falls below 200 cells/mm^3 [42]. It has been noted that HIV patients, which develop the disease PCP are most often already in the AIDS stage of the disease progression and have had failure or non-compliance in their HAART regimes [66].

In non-HIV patients, PCP is usually a more serious public health concern, with faster onset and progression of the disease and symptoms [83]. The reason for this is likely due to underlying medical conditions, which may have more serious disease presentations than HIV, and also some non-HIV patients being less aware they're at risk to PCP [79, 104]. This often leads to no preventative measures implemented, and a treatment delay [117]. Non-HIV patients with lowered CD4+ T cell levels are

highest at risk of infection [80]. Among these, individuals suffering from hematological neoplasms, in particular patients with leukaemia and lymphoma, have the highest risk developing PCP infections, due to immunosuppressive chemotherapies or long-term steroid use [105, 118]. This risk is further heightened in patients with acute lymphoblastic leukaemia [43]. Conversely, although chances of infection are lower, the mortality rate is highest among patients with solid tumours [81, 119]. An increasing risk of PCP is also developing within solid organ transplant recipients [9]. Without of prophylaxis, reported attack rates of PCP are up to 10%, 41% and 43% in renal, heart and lung transplant recipients, respectively [5]. PCP will often manifest 3 to6 months after transplantation, but risk is significantly minimized with compliance to prophylaxis.

The most common treatment related risk for PCP is receiving corticosteroids, cytotoxic therapies, or immunosuppressive agents, in particularly cyclosproine A, antithymocyte globulin, and glucocorticosteroids [82].

1.12 Genotyping

An effective population genetics study of *Pneumocystis* has been stunted due to its inability to be isolated, grown and sustained in pure culture [7]. Over the past two decades, molecular and genotyping techniques have helped to better understand *Pneumocystis*. Genotyping has helped to understand the prevalence, transmission, guide treatment regimes, detect drug resistance, virulence, cell biology and pathogenesis of *Pneumocystis* [83]. Epidemiological studies have truly been advanced with the use of genotyping tools. They have provided information about the genetic diversity of *Pneumocystis*, differences between species and genotypes, clinical outcomes, geographic distribution, modes of transmission, and population analysis [84].

For *Pneumocystis* genotyping studies, techniques used range from simple karyotyping [85], multilocus enzyme electrophoresis (MLEE) [86], restriction fragment length polymorphism (RFLP)

[87, 88], type-specific oligonucleotide hybridization (TSOH) (also known as allele-specific OH or dot blotting) [89, 90], single strand conformation polymorphism (SSCP) [91], short tandem repeat (STR) Typing (also known as microsatellites or simple sequence repeats (SSRs)) [92], multilocus sequence typing (MLST) [93], and finally, to whole genome sequencing (WGS) [94]. Each technique has its own advantages and disadvantages, each developed and improved from the technique before, increasing our knowledge and helping widen *Pneumocystis* studies.

Table 1.2. Description of all techniques used for genotyping within the literature *Pneumocystis* spp. References noted above with respective techniques.

Techniques	Description	Advantages	Disadvantages
Karyotyping	Chromosomes of an organism separated by pulse-field gel electrophoresis.	Relatively fast and simple and can assess many samples at one time.	Can only be applied to *Pneumocystis* spp. that can be propagate in animal models and detects low levels of genetic discrimination.
MLEE	Differentiation of an organism's isoenzymes by assessing their migration in a non-denaturing gel.	Previously gold standard for phylogeny by analysing the patterns of different samples.	Highly fastidious, time consuming and need a large number of cells, making it more adapt to only animal models. Can only detect changes when they result in amino acid change.
TSHO	Uses PCR to amplify DNA to test for already known mutations, with the use of two probes (radioactive, enzymatic, or fluorescent). One probe is for wild-types and the second is corresponding to the mutation.	Can detect genetic polymorphisms quicker than direct DNA sequencing and can be used for routine screening.	Can only be used to test against known allele variations.
RFLP	Detects variations in homologous DNA molecules. Restriction enzymes are used to digest DNA, and fragments are separated by length. (single-point mutations	RFLP loci are located throughout the entire genome, allowing for multiple samples to be analysed at once, as well as being highly reproducible.	RFLP combines gel electrophoresis, capillary transfer of fragments into a membrane, followed by southern hybridization, which is time consuming, expensive and requires high quality DNA.
SSCP	Detects variants by identifying single nucleotide polymorphisms via measuring differences through electrophoretic mobility. Noted variants have a different mobility compared to wild type samples.	Inexpensive, quick, and simple to use as it does not require sequencing. Can be purchased from easily available commercial kits to allow quick screening of samples	4 entirely (ITS1, *26S*, *mt26S*, *DHPS*) and 1 partial loci (*β-TUB*) have been used for SSCP in *P. jirovecii* typing, it is no longer a popular typing method due to poor throughput and poor migration quality.

STR-Typing	Detects fragments of DNA with core repeating sections, often 2-7 nucleotides long, which are then tandemly repeated 6 – 12 times. Often 8 – 12 STR's selected in a typing scheme. PCR amplicons are separated by capillary electrophoresis and then typed by comparison to a ladder.	Inexpensive, quick, and simple to use as it does not require sequencing. Reported discriminatory power reached in this study was 0.999.	Relatively low typing completed to date, with no standardised scheme or evaluation of STRs for use. There has neem no database established, making it extremely difficult for hospital and laboratory comparison and collaboration.
MLST	Using PCR to amplify 3-8 genetic loci of a sample, then comparing sequences and classifying mutations into different allele and sequence types.	Currently the gold standard. Highly reproducible, relatively accessible and with high throughput. Depending on alleles, can have very high discriminatory power.	Weak detection mixed infections, no established consensus scheme, making it difficult for hospital and laboratory comparison and collaboration if they do not use the same scheme.
WGS	Can use a variety of techniques to sequence the entire genome of a sample.	Highly specific and the most discriminatory and sequences every nucleotide in the genome.	WGS is highly costly, time consuming, laborious and complex, requiring analysis from bioinformaticians.

Regardless of recent advancements in molecular technology, MLST remains the gold standard for genotyping of *Pneumocystis* strains [118, 120, 121, 122]. In general, MLST involves the comparison of a set of sequences of housekeeping genes between different strains or genotypes. These genetic fragments, which are highly discriminative, are amplified via PCR, sequenced and the sequences are then compared between strains for molecular epidemiologic applications. When analysing and comparing, any unique sequences are given an allele number and then combined into an allelic profile, which is assigned its own sequence type. Degrees of relatedness are investigated by comparing the allelic profiles [69]. A large number of studies in different pathogenic fungi show that MLST genotyping techniques to have a high discriminatory power, great sensitivity, reproducible, and have the possibility of exchanging data between different laboratories. Targeting a multicopy gene, or using nested-PCR, which increases the locus detection rate, can furtherimprove the sensitivity [83].

Currently there are 17 DNA regions which are used in different MLST schemes to study the allelic polymorphisms of *P. jirovecii* isolates. They include: *mitochondrial rRNA* gene (*mt26S*), internal transcribed spacer 1 (ITS1), internal transcribed spacer 2 (ITS2*), β-tubulin (β-TUB*), *large subunit of the rRNA* gene (*26S rRNA*) (as well as *5.8S*, *18S*, *23S*), *mitochondrial small subunit* (*mtSSU*) *rRNA, superoxide dismutase (SOD), cytochrome b (CYB*), *thymidylate synthase (TS), 5.8S rRNA*, *AROM* protein, *thioredoxin reductase (TRR1), upstream conserved sequence (UCS),* m*ajor surface glycoprotein (MSG), serine endopeptidase (KEX1), dihydrofolate reductase* (*DHFR*), and *dihydropteroate synthase (DHPS)* [89, 93].

In 1997, P. Hauser and his team first suggested a four-locus-based gene scheme, using ITS1, 26S, *mt26S*, and *β-TUB*. This MLST scheme was able to differentiate between genotypes and investigate *P. jirovecii* diversity with a high discriminatory power [95]. Since then, there have been dozens of combinations of genetic loci used in various MSLT schemes, varying between countries and laboratories, using between 2 to 8 loci reviewed by Esteves and Matos, 2010, [89]. Till today there is no consensus MLST scheme among research teams.

In 2013, a French team produced a MLST scheme, using conventional single-round PCR, which would be most effective in determining epidemiological variability. The study showed that the loci of choice are crucial when analysing genetic variation, with a minimum of three to four locinecessary for a satisfactory discrimination between strains. MLST schemes with insufficient performance can lead to isolates being wrongly identified as identical or different. From this study, two MLST schemes were proposed to be used in future *Pneumocystis* studies, which would yieldthe strongest discriminatory power. They were Hauser's original scheme (ITS1, *26S, mt26S* and *β-TUB*), with a H index of 0.0987, and a new simpler scheme, consisting of 3 genetic loci, *SOD, mt26S,* and *CYB*, and yielding a H value of 0.996 [93].

Another team studied the epidemiological links of a *Pneumocystis* outbreak in Australia, and the detection of *Pneumocystis* in young children without PCP, using again a different MLST

scheme. This MLST scheme employed four genetic loci, ITS1/2, *β-TUB, mt26S,* and *DHPS* (Table 1.3), which also yielded a satisfactory discriminatory power to distinguish unrelated isolates from nosocomial outbreak isolates [96].

Table 1.3. Genetic loci used in Australian MLST scheme. Loci used, with respective forward and reverse primers, product size, and GenBank reference numbers. Table compiled from *Chen et al.,* 2014.[96]

MLST Locus	Nucleotide sequence of primers	Product size (bp)	GenBank reference accession number
β -TUB	Forward primer: BTubulinF 5´-TCATTAGGTGGTGGAACGGG-3´ Reverse primer: BTubulinR: 5´-ATCACCATATCCTGGATCCG-3´	309	EU979560
DHPS	Forward primer: DHPS-3 5´-GCGCCTACACATATTATGGCCATTTTAAATC-3´ Reverse primer: DHPS-4 5´-GGAACTTTCAACTTGGCAACCAC-3´	371	AJ586567
mt26S	Forward primer: mt26SF 5´-GATGGCTGTTTCCAAGCCCA-3´ Reverse primer: mt26SR 5´-GTGTACGTTGCAAAGTACTC-3´	307	M58605
ITS	First round Forward primer: 1724F2 5´-AGTTGATCAAATTTGGTCATTTAGAG-3´ Reverse primer: ITS2R 5´-CTCGGACGAGGATCCTCGCC-3´ Second round Forward primer: ITS1F2 5´-CGTAGGTGAACCTGCGGAAGGATC-3´ Reverse primer: ITS2R1 5´-GTTCAGCGGGTGATCCTGCCTG-3´	549	U07220

The four genetic loci used in the Australian MLST scheme were chosen due to their high discriminative power. The ITS1 sequence is found within nuclear rRNA operon between the genes of the *18S rRNA* and the *5.8S rRNA,* and ITS2 is located between the genes of the *5.8S rRNA* and the *26S rRNA*. This area shows a high level of polymorphism, which is used for allelic typing [97]. *mt26S* locus has a high degree of genetic conservation and is hence useful for distinguishing

variations. Polymorphisms at this locus are often seen at nucleotide positions 85 and 248 [77]. Sulfonamide antimicrobial drugs target the *DHPS* gene, creating polymorphisms at codons 55 and 57. Analysis of *P. jirovecii* strains with mutations in this locus can assess both the genotyping, and the prophylaxis and treatment failure [98]. *β-TUB* is a highly conserved single-copy gene, and there is 70%-90% identity at the amino acid level between *β-TUB* sequences from widely divergent organisms. This makes it a useful tool for typing and assessing polymorphisms between strains of *Pneumocystis* [99]. The generated data were made available via the ISHAM fungal MLST webpage: mlst.mycologylab.org. Since then, two other Australian publications have been produced using the same MLST scheme, but no other submissions have been made for genotyping reports of *Pneumocystis* in Australia or using that ISHAM MLST scheme [100, 101].

The existence of different MLST typing schemes, on one side, has contributed to the general understanding of the epidemiology of *Pneumocystis,* however, one the other side, it makes it very hard to compare different typing results. In addition, to due the nature of the *Pneumocystis* life cycle and its inability to be cultured *in vitro,* there is still variable amplification rates between even the more robust MLST schemes [89, 93]. This can be explained by: 1) low fungal burden in some patients, in particular colonized patients, 2) differences in the yield of the extraction methods used for respiratory samples, 3) the purity of the samples, and the percentage of mixture with other disease agents within the sample [83].

Further development of the currently used genotyping techniques is necessary in order to expand and deepen the understanding of the study gaps present in *Pneumocystis* research.

Especially, the possibility of further fine tuning, and tweaking of the genotyping techniques, and their conditions and protocols should be explored. Improvement in the genotyping and epidemiological advances, and the possible development of a world-wide consensus genotyping scheme can provide a more efficient, unified typing method, that will allow for the rapid analysis of *Pneumocystis* isolates, especially in primary investigations of PCP.

1.13 Aims

The overall aim of this PhD thesis is to contribute to a better understanding of the genetic diversity and epidemiology of *Pneumocystis jirovecii.*

The specific aims are:

Aim 1: To compare current laboratory protocols used for identification and genotyping of human *Pneumocystis* samples and establish ways to improve and streamline DNA collection. (Chapter 2)

Aim 2: To identify pitfalls and optimize the current *P. jirovecii* ISHAM MLST Scheme, as well as analyzing its efficacy when used to genotype a PCP cohort. (Chapter 3)

Aim 3: To develop and apply a global consensus *P. jirovecii* MLST Scheme to PCP samples. (Chapter 4)

Aim 4: The conduct an epidemiological analysis of global *P. jirovecii* isolates using the newly developed consensus MLST scheme. (Chapter 5)

Aim 5: To assess the effectiveness of the *DHPS* locus in diagnosis and genotyping of *P. jirovecii.* (Chapter 6)

Aim 6: To analyze the ability of next generation sequencing for the identification and genotyping typing of *Pneumocystis* spp. (Chapter 7)

1.14 References

1. Chagas, C., *Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem.* Memórias do Instituto Oswaldo Cruz, 1909. **1**: p. 159-218.

2. Redhead, S.A., *et al.*, *Pneumocystis and Trypanosoma cruzi: Nomenclature and Typifications.* Journal of Eukaryotic Microbiology, 2006. **53**(1): p. 2-11.

3. Calderón Sandubete, E., Y. de Armas Rodríguez, and V. Capó de Paz, *Pneumocystis jirovecii: one hundred years of history.* Revista cubana de medicina tropical, 2011. **63**(2): p. 97.

4. Dutz, W., *Pneumocystis carinii pneumonia.* Pathology annual, 1970. **5**: p. 309.

5. Sepkowitz, K.A., A.E. Brown, and D. Armstrong, *Pneumocystis carinii pneumonia without acquired immunodeficiency syndrome: more patients, same risk.* Archives of internal medicine, 1995. **155**(11): p. 1125-1128.

6. Frenkel, J., *Pneumocystis jiroveci n. sp. from man: morphology, physiology, and immunology in relation to pathology.* National Cancer Institute Monograph, 1976. **43**: p. 13-30.

7. Krajicek, B.J., C.F. Thomas, and A.H. Limper, *Pneumocystis pneumonia: current concepts in pathogenesis, diagnosis, and treatment.* Clinics in Chest Medicine, 2009. **30**(2): p. 265-278.

8. Gottlieb, M.S., *et al.*, *Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency.* New England Journal of Medicine, 1981. **305**(24): p. 1425-1431.

9. Catherinot, E., *et al.*, *Pneumocystis jirovecii Pneumonia.* Infectious disease clinics ofNorth America, 2010. **24**(1): p. 107-138.

10. San-Andrés, F.-J., *et al.*, *Incidence of acquired immunodeficiency syndrome-associated opportunistic diseases and the effect of treatment on a cohort of 1115 patients infected with human immunodeficiency virus, 1989–1997.* Clinical infectious diseases, 2003. **36**(9): p. 1177-1185.

11. Stringer, S.L., *et al.*, *Pneumocystis carinii: sequence from ribosomal RNA implies a close relationship with fungi.* Experimental parasitology, 1989. **68**(4): p. 450-461.

12. Limper, A.H., *Pneumocystis nomenclature.* Clinical Infectious Diseases, 2006. **42**(8): p. 1210-1211.

13. Thomas Jr, C.F. and A.H. Limper, *Pneumocystis pneumonia.* New England Journal of Medicine, 2004. **350**(24): p. 2487-2498.

14. Stringer, J.R., M. Cushion, and A.E. Wakefield, *New Nomenclature for the Genus Pneumocystis.* Journal of Eukaryotic Microbiology, 2001. **48**(s1): p. 184s-189s.

15. Gigliotti, F., *Pneumocystis carinii: has the name really been changed?* Clinical infectious diseases, 2005. **41**(12): p. 1752-1755.

16. Cushion, M.T. and J.R. Stringer, *Has the name really been changed? It has for most researchers.* Clinical Infectious Diseases, 2005. **41**(12): p. 1756-1758.

17. Hauser, P.M., *Is the unique camouflage strategy of Pneumocystis associated with its particular niche within host lungs?* PLoS pathogens, 2019. **15**(1): p. e1007480.

18. Gigliotti, F., *et al.*, *Pneumocystis carinii is not universally transmissible between mammalian species.* Infection and immunity, 1993. **61**(7): p. 2886-2890.

19. Edman, J.C., *et al.*, *Ribosomal RNA sequence shows Pneumocystis carinii to be a member of the fungi.* Nature, 1988. **334**(6182): p. 519-522.

20. Sanches, E.M.C., *et al.*, *Phylogenetic analysis of Pneumocystis from pig lungs obtained from slaughterhouses in southern and midwestern regions of Brazil.* Arquivo Brasileiro de Medicina Veterinária e Zootecnia, 2011. **63**: p. 1154-1159.

21. Keely, S.P., *et al.*, *Phylogenetic identification of Pneumocystis murina sp. nov., a new species in laboratory mice.* Microbiology, 2004. **150**(5): p. 1153-1165.

22. Chabé, M., *et al.*, *Pneumocystis carinii and Pneumocystis wakefieldiae in wild Rattus norvegicus trapped in Thailand.* Journal of Eukaryotic Microbiology, 2010. **57**(2): p. 213-217.

23. Babb-Biernacki, S.J., J.A. Esselstyn, and V.P. Doyle, *Rethinking host range in Pneumocystis.* PLOS Pathogens, 2020. **16**(9): p. e1008824.

24. Ma, L., O.H. Cissé, and J.A. Kovacs, *A molecular window into the biology and epidemiology of Pneumocystis spp.* Clinical microbiology reviews, 2018. **31**(3): p. e00009-18.

25. Latinne, A., *et al.*, *Genetic diversity and evolution of Pneumocystis fungi infecting wild Southeast Asian murid rodents.* Parasitology, 2018. **145**(7): p. 885-900.

26. Thomas, C.F. and A.H. Limper, *Current insights into the biology and pathogenesis of Pneumocystis pneumonia.* Nature Reviews Microbiology, 2007. **5**(4): p. 298-308.

27. Sloand, E., *et al.*, *The Challenge of Pneumocystis carinii Culture1.* Journal of Eukaryotic Microbiology, 1993. **40**(2): p. 188-195.

28. Schildgen, V., *et al.*, *Pneumocystis jirovecii can be productively cultured in differentiated CuFi-8 airway cells.* MBio, 2014. **5**(3): p. e01186-14.

29. Liu, Y., G.A. Fahle, and J.A. Kovacs, *Inability to culture Pneumocystis jirovecii.* MBio, 2018. **9**(3): p. e00939-18.

30. Huang, L., *et al.*, *An Official ATS Workshop Summary: Recent advances and future directions in Pneumocystis pneumonia (PCP).* Proceedings of the American Thoracic Society, 2006. **3**(8): p. 655-664.

31. Matsumoto, Y. and Y. Yoshida, *Sporogony in Pneumocystis carinii: Synaptonemal Complexes and Meiotic Nuclear Divisions Observed in Precysts1.* The Journal of protozoology, 1984. **31**(3): p. 420-428.

32. Krajicek, B.J., A.H. Limper, and J.C.F. Thomas, *Advances in the biology, pathogenesis and identification of Pneumocystis pneumonia.* Current Opinion in Pulmonary Medicine, 2008. **14**(3): p. 228-234.

33. Neese, L.W., *et al.*, *Vitronectin, fibronectin, and gp120 antibody enhance macrophage release of TNF-alpha in response to Pneumocystis carinii.* The Journal of Immunology, 1994. **152**(9): p. 4549-4556.

34. Limper, A. and W. Martin 2nd, *Pneumocystis carinii: inhibition of lung cell growth mediated by parasite attachment.* Journal of Clinical Investigation, 1990. **85**(2): p. 391.

35. Limper, A.H., J.S. Hoyte, and J.E. Standing, *The role of alveolar macrophages in Pneumocystis carinii degradation and clearance from the lung.* The Journal of clinical investigation, 1997. **99**(9): p. 2110.

36. Beck, J.M., *et al.*, *Host Defenses Against Pneumocystis carinii in Mice Selectively Depleted of CD4+ Lymphocytes.* Chest, 1993. **103**: p. 116S-118S.

37. Harmsen, A.G. and M. Stankiewicz, *Requirement for CD4+ cells in resistance to Pneumocystis carinii pneumonia in mice.* The Journal of experimental medicine, 1990. **172**(3): p. 937-945.

38. Azoulay, E., *et al.*, *AIDS-related Pneumocystis carinii pneumonia in the era of adjunctive steroids: implication of BAL neutrophilia.* American journal of respiratory and critical care medicine, 1999. **160**(2): p. 493-499.

39. Limper, A.H., *et al.*, *Pneumocystis carinii pneumonia: differences in lung parasite number and inflammation in patients with and without AIDS.* American review of respiratory disease, 1989. **140**(5): p. 1204-1209.

40. Barry, S. and M. Johnson, *Pneumocystis carinii pneumonia: a review of current issues in diagnosis and management.* HIV medicine, 2001. **2**(2): p. 123-132.

41. Walzer, P.D., *et al.*, *Pneumocystis carinii pneumonia in the United States: epidemiologic, diagnostic, and clinical features.* Annals of internal medicine, 1974. **80**(1): p. 83-93.

42. Phair, J., *et al.*, *The risk of Pneumocystis carinii pneumonia among men infected with human immunodeficiency virus type 1.* New England Journal of Medicine, 1990. **322**(3): p. 161-165.

43. Russian, D.A. and S.J. Levine, *Pneumocystis carinii pneumonia in patients without HIV infection.* The American journal of the medical sciences, 2001. **321**(1): p. 56-65.

44. Kovacs, J.A., *et al.*, *Pneumocystis carinii pneumonia: a comparison between patients with the acquired immunodeficiency syndrome and patients with other immunodeficiencies.* Annals of internal medicine, 1984. **100**(5): p. 663-671.

45. DeLorenzo, L., *et al.*, *Roentgenographic patterns of Pneumocystis carinii pneumonia in 104 patients with AIDS.* Chest Journal, 1987. **91**(3): p. 323-327.

46. Kennedy, C.A. and M.B. Goetz, *Atypical roentgenographic manifestations of Pneumocystis carinii pneumonia.* Archives of internal medicine, 1992. **152**(7): p. 1390- 1398.

47. Medscape. *Progressive Respiratory Disturbances in HIV+ Man*. 2012 [cited 2015; Available from: http://reference.medscape.com/features/slideshow/resp-hiv/slide18.

48. Gigliotti, F., A. Harmsen, and T. Wright, *Characterization of transmission of Pneumocystis carinii f. sp. muris through immunocompetent BALB/c mice.* Infection and immunity, 2003. **71**(7): p. 3852-3856.

49. Larsen, H.H., *et al.*, *Primary Pneumocystis infection in infants hospitalized with acute respiratory tract infection.* Emerging infectious diseases, 2007. **13**(1): p. 66.

50. Ribes, J.A., *et al.*, *PCR detection of Pneumocystis carinii in bronchoalveolar lavage specimens: analysis of sensitivity and specificity.* Journal of clinical microbiology, 1997. **35**(4): p. 830-835.

51. Procop, G., *et al.*, *Detection of Pneumocystis jiroveci in respiratory specimens by four staining methods.* Journal of clinical microbiology, 2005. **43**(4): p. 2038.

52. Kovacs, J.A., *et al.*, *Diagnosis of Pneumocystis carinii pneumonia: improved detection in sputum with use of monoclonal antibodies.* New England Journal of Medicine, 1988. **318**(10): p. 589-593.

53. Azoulay, E., *et al.*, *Polymerase chain reaction for diagnosing Pneumocystis pneumonia in non-HIV immunocompromised patients with pulmonary infiltrates.* Chest Journal, 2009. **135**(3): p. 655-661.

54. Wakefield, A., *et al.*, *Detection of Pneumocystis carinii with DNA amplification.* The Lancet, 1990. **336**(8713): p. 451-453.

55. Flori, P., *et al.*, *Comparison between real-time PCR, conventional PCR and different staining techniques for diagnosing Pneumocystis jiroveci pneumonia from bronchoalveolar lavage specimens.* Journal of medical microbiology, 2004. **53**(7): p.603-607.

56. Fujisawa, T., *et al.*, *Real-time PCR is more specific than conventional PCR for induced sputum diagnosis of Pneumocystis pneumonia in immunocompromised patients without HIV infection.* Respirology, 2009. **14**(2): p. 203-209.

57. Huggett, J.F., *et al.*, *Development and evaluation of a real-time PCR assay for detection of Pneumocystis jirovecii DNA in bronchoalveolar lavage fluid of HIV-infected patients.* Thorax, 2008. **63**(2): p. 154-159.

58. Otahbachi, M., K. Nugent, and D. Buscemi, *Granulomatous Pneumocystis jiroveci Pneumonia in a patient with chronic lymphocytic leukemia: a literature review and hypothesis on pathogenesis.* The American journal of the medical sciences, 2007. **333**(2): p. 131-135.

59. Sattler, F.R., *et al.*, *Trimethoprim-sulfamethoxazole compared with pentamidine for treatment of Pneumocystis carinii pneumonia in the acquired immunodeficiency*

syndrome: a prospective, noncrossover study. Annals of internal medicine, 1988. **109**(4): p. 280-287.

60. Bucher, H.C., *et al.*, *Meta-analysis of prophylactic treatments against Pneumocystis carinii pneumonia and toxoplasma encephalitis in HIV-infected patients.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 1997. **15**(2): p. 104-114.

61. Stein, D., *et al.*, *Use of low-dose trimethoprim-sulfamethoxazole thrice weekly for primary and secondary prophylaxis of Pneumocystis carinii pneumonia in human immunodeficiency virus-infected patients.* Antimicrobial agents and chemotherapy, 1991. **35**(9): p. 1705-1709.

62. Dumoulin, A., *et al.*, *Transmission of Pneumocystis carinii disease from immunocompetent contacts of infected hosts to susceptible hosts.* European Journal of Clinical Microbiology and Infectious Diseases, 2000. **19**(9): p. 671-678.

63. Nahimana, A., *et al.*, *Association between a specific Pneumocystis jiroveci dihydropteroate synthase mutation and failure of pyrimethamine/sulfadoxine prophylaxis in human immunodeficiency virus–positive and–negative patients.* The Journal of infectious diseases, 2003. **188**(7): p. 1017-1023.

64. Iliades, P., S.R. Meshnick, and I.G. Macreadie, *Analysis of Pneumocystis jirovecii DHPS alleles implicated in sulfamethoxazole resistance using an Escherichia coli model system.* Microbial Drug Resistance, 2005. **11**(1): p. 1-8.

65. Iliades, P., S.R. Meshnick, and I.G. Macreadie, *Mutations in the Pneumocystis jirovecii DHPS gene confer cross-resistance to sulfa drugs.* Antimicrobial agents and chemotherapy, 2005. **49**(2): p. 741.

66. Walzer, P.D., *et al.*, *Early predictors of mortality from Pneumocystis jirovecii pneumonia in HIV-infected patients: 1985–2006.* Clinical infectious diseases, 2008. **46**(4): p. 625- 633.

67. Pagano, L., *et al.*, *Pneumocystis carinii pneumonia in patients with malignant haematological diseases: 10 years' experience of infection in GIMEMA centres.* British journal of haematology, 2002. **117**(2): p. 379-386.

68. Morris, A., C.B. Beard, and L. Huang, *Update on the epidemiology and transmission of Pneumocystis carinii.* Microbes and infection, 2002. **4**(1): p. 95-103.

69. Beard, C.B., *et al.*, *Genetic variation in Pneumocystis carinii isolates from different geographic regions: implications for transmission.* Emerging infectious diseases, 2000. **6**(3): p. 265.

70. Garvy, B.A. and M.H. Qureshi, *Delayed inflammatory response to Pneumocystis carinii infection in neonatal mice is due to an inadequate lung environment.* The Journal of Immunology, 2000. **165**(11): p. 6480-6486.

71. Chen, W., F. Gigliotti, and A. Harmsen, *Latency is not an inevitable outcome of infection with Pneumocystis carinii.* Infection and immunity, 1993. **61**(12): p. 5406-5409.

72. An, C.L., F. Gigliotti, and A.G. Harmsen, *Exposure of immunocompetent adult mice to Pneumocystis carinii f. sp. muris by cohousing: growth of P. carinii f. sp. muris and host immune response.* Infection and immunity, 2003. **71**(4): p. 2065-2070.

73. Phipps, L.M., *et al.*, *Nosocomial Pneumocystis jirovecii pneumonia: lessons from a cluster in kidney transplant recipients.* Transplantation, 2011. **92**(12): p. 1327-1334.

74. Olsson, M., *et al.*, *Genotypes of clustered cases of Pneumocystis carinii pneumonia.* Scandinavian journal of infectious diseases, 2001. **33**(4): p. 285-289.

75. Fritzsche, C., *et al.*, *Pneumocystis jirovecii colonization among renal transplantrecipients.* Nephrology, 2013. **18**(5): p. 382-387.

76. Vargas, S.L., *et al.*, *Search for primary infection by Pneumocystis carinii in a cohort of normal, healthy infants.* Clinical infectious diseases, 2001. **32**(6): p. 855-861.

77. Medrano, F.J., *et al.*, *Pneumocystis jirovecii in general population.* Emerging Infectious Diseases, 2005. **11**(2): p. 245-250.

78. Morris, A., *et al.*, *Prevalence and clinical predictors of Pneumocystis colonization among HIV-infected men.* Aids, 2004. **18**(5): p. 793-798.

79. Roblot, F., *et al.*, *Analysis of underlying diseases and prognosis factors associated with Pneumocystis carinii pneumonia in immunocompromised HIV-negative patients.* European journal of clinical microbiology and infectious diseases, 2002. **21**(7): p. 523-531.

80. Monnet, X., *et al.*, *Critical care management and outcome of severe Pneumocystis pneumonia in patients with and without HIV infection.* Crit Care, 2008. **12**(1): p. R28.

81. Kofteridis, D.P., *et al.*, *Predisposing factors, clinical characteristics and outcome of Pneumonocystis jirovecii pneumonia in HIV-negative patients.* Journal of Infection and Chemotherapy, 2014. **20**(7): p. 412-416.

82. de Boer, M.G., *et al.*, *Risk factors for Pneumocystis jirovecii pneumonia in kidney transplant recipients and appraisal of strategies for selective use of chemoprophylaxis.* Transplant Infectious Disease, 2011. **13**(6): p. 559-569.

83. Sokulska, M., *et al.*, *Pneumocystis jirovecii—from a commensal to pathogen: clinical and diagnostic review.* Parasitology research, 2015. **114**(10): p. 3577-3585.

84. Monroy-Vaca, E.X., *et al.*, *Genetic diversity of Pneumocystis jirovecii in colonized Cuban infants and toddlers.* Infection, genetics and evolution, 2014. **22**: p. 60-66.

85. Hong, S.-T., *et al.*, *Pneumocystis carinii karyotypes.* Journal of Clinical Microbiology, 1990. **28**(8): p. 1785-1795.

86. Hauser, P.M., *et al.*, *III. Typing methods to approach Pneumocystis carinii genetic heterogeneity.* FEMS Immunology & Medical Microbiology, 1998. **22**(1-2): p. 27-35.

87. Costa, M.C., *et al.*, *Dihydropteroate Synthase (DHPS) Genotyping by PCR-RFLP Analysis of Pneumocystis jirovecii Repeated Isolates from HIV-Infected Patients: A Preliminary Study.* The Journal of Eukaryotic Microbiology, 2003. **50**(6): p. 607-608.

88. Dai, S. and Y. Long, *Genotyping Analysis Using an RFLP Assay*, in *Plant Genotyping:Methods and Protocols*, J. Batley, Editor. 2015, Springer New York: New York, NY. p. 91- 99.

89. Esteves, F., *et al.*, *Pneumocystis jirovecii multilocus genotyping profiles in patients from Portugal and Spain.* Clinical Microbiology and Infection, 2008. **14**(4): p. 356-362.

90. Bertorini, T.E., *2 - Neurologic Evaluation and Ancillary Tests*, in *Neuromuscular Case Studies*, T.E. Bertorini, Editor. 2008, Butterworth-Heinemann: Philadelphia. p. 27-76.

91. Mathis, A., *et al.*, *Reliable one-tube nested PCR for detection and SSCP-typing of Pneumocystis carinii.* Journal of Eukaryotic Microbiology, 1996. **43**(5): p. 7-S7.

92. Alanio, A., *et al.*, *European study on Pneumocystis jirovecii short tandem repeats genotyping reveals wide population diversity with geographic specificities.* Journal de Mycologie Médicale, 2016. **26**(2): p. e5-e6.

93. Maitte, C., *et al.*, *Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* Journal of clinical microbiology, 2013. **51**(9): p. 2843-2849.

94. Sesterhenn, T.M., *et al.*, *Generation of Sequencing Libraries for the Pneumocystis Genome Project.* The Journal of Eukaryotic Microbiology, 2003. **50**(6): p. 663-665.

95. Hauser, P.M., Patrick Francioli, Jacques Bille, Amalio Telenti, and Dominique S. Blanc., *Typing of Pneumocystis carinii f. sp. hominis by single-strand conformationpolymorphism of four genomic regions.* Journal of clinical microbiology 1997. **35**(12): p. 3086-3091.

96. Chen, S., Brian Nankivell, Carolina Firacative, Kathy Kable, Debbie Marriott, Peter MacDonald, Wieland Meyer, and Jeremy Chapman. , *Hospital-acquired Pneumocystis pneumonia: a renewed concern?* Microbiology Australia, 2014. 35(1): p. 57-59.

97. Lu, J.-J., *et al.*, *Typing of Pneumocystis carinii strains that infect humans based on nucleotide sequence variations of internal transcribed spacers of rRNA genes.* Journal of clinical microbiology, 1994. 32(12): p. 2904-2912.

98. Tyagi, A., *et al.*, *Study of dihydropteroate synthase (DHPS) gene mutations among isolates of Pneumocystis jiroveci.* Indian Journal of Medical Research, 2008. 128(6): p. 734.

99. Edlind, T.D., M.S. Bartlett, and J.W. Smith, *Characterization of the beta-tubulin gene of Pneumocystis carinii.* The Journal of protozoology, 1991. 38(6): p. 62S.

100. Nankivell, B.J., *et al.*, *Molecular epidemiology linking multihospital clusters of opportunistic Pneumocystis jirovecii pneumonia.* Clinical infectious diseases, 2013. **57**(7): p. 1058-1059.

101. Chua, K.Y., *et al.*, *Colonisation with Pneumocystis jirovecii in Australian infants.* Pathology-Journal of the RCPA, 2015. 47(5): p. 489-490.

102. Afessa, B., Green, W., Chiao, J. & Frederick, W. *Pulmonary complications of HIV infection: autopsy findings*, 1998. Chest 113, 1225–1229

103. Hsu, J.M., Hass, A., Gingras, MA. *et al. Radiographic features in investigated for Pneumocystis jirovecii pneumonia: a nested case-control study.* BMC Infect Dis 2020. 20, 492

104. Calderón, E. J., *et al*. "*Epidemiology of Pneumocystis carinii pneumonia in southern Spain."* Clinical microbiology and infection 2004. 10.7: 673-676.

105. Roux, Antoine, et al. "*Pneumocystis jirovecii pneumonia in patients with or without AIDS, France*." Emerging infectious diseases 2014. 20.9 : 1490.

106. Cushion, Melanie, T. *Are Members of the Fungal Genus Pneumocystis (a) Commensals; (b) Opportunists; (c) Pathogens; or (d) All of the Above?* PLoS Pathog 2010. 6(9): e1001009.

107. Montes-Cano, M.A., *et al.*, *Vertical transmission of Pneumocystis jirovecii in humans.* Emerging infectious diseases, 2009. 15(1): p. 125

108. Szydłowicz, Magdalena*, et al. Pneumocystis jirovecii colonization in preterm newborns with respiratory distress syndrome.* The Journal of Infectious Diseases 2021

109. Huang, Laurence, *et al. Dihydropteroate synthase gene mutations in Pneumocystis and sulfa* resistance. Emerging infectious diseases 2004. 10.10: 1721.

110. Lee, Sang Min, *et al. A case of pneumonia caused by Pneumocystis jirovecii resistant to trimethoprim-sulfamethoxazole.* The Korean Journal of Parasitology, 2015. 53.3: 321.

111. Kandil, O. *et al. Human immunodeficiency virus type 1 infection of human macrophages modulates the cytokine response to Pneumocystis carinii.* Infection and Immunity, 1994. 62:644-650

112. Campbell JH, *et al. The importance of monocytes and macrophages in HIV pathogenesis, treatment, and cure*. AIDS, 2014; 28 :2175-2187

113. Zhang, Jianmin, *et al*. *Pneumocystis activates human alveolar macrophage NF-κB signaling through mannose receptors.* Infection and Immunity. 2004. 72.6 3147-3160.

114. Segal, Leopoldo N., *et al*. *HIV-1 and bacterial pneumonia in the era of antiretroviral therapy.* Proceedings of the American Thoracic Society, 2011. 8.3 : 282-287

115. Pereira-Díaz, Estefanía, *et al. Changing trends in the epidemiology and risk factors of Pneumocystis pneumonia in Spain*. Frontiers in public health, 2019. 275

116. Shaddock, Erica. *Pneumocystis jirovecii: A Fresh Look.* Clinical Pulmonary Medicine, 2016. 23.1: 37-43

117. Thomas, Sherine, *et al. Increasing frequency of Pneumocystis jirovecii pneumonia in renal transplant recipients in the United Kingdom: clonal variability, clusters, and geographic location.* Clinical infectious diseases, 2011. 53.3: 307-308.

118. Veronese, Giacomo, *et al. Single-center outbreak of Pneumocystis jirovecii pneumonia in heart transplant recipients.* Transplant Infectious Disease, 2018. 20.3: e12880

119. Roux, Antoine, *et al. Pneumocystis jirovecii pneumonia in patients with or without AIDS,*

France. Emerging infectious diseases, 2014. 20.9: 1490

120. Desoubeaux G *et al. Epidemiological outbreaks of Pneumocystis jirovecii pneumonia are not limited to kidney transplant recipients: genotyping confirms common source of transmission in a liver transplantation unit.* J Clin Microbiol. 2016; 54: 1314–1320.

121. Szydłowicz, M., *et al., Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients—preliminary report.* Parasitology research, 2019. 118(1): p. 181-189.

122. Depypere, M., V. Saegeman, and K. Lagrou, *Typing of Pneumocystis jirovecii by multilocus sequencing: evidence of outbreak?* European Journal of Clinical Microbiology & Infectious Diseases, 2016. 35(6): p. 911-916

Chapter 2

Optimization of Lab Protocols

2.1 Introduction

P. jirovecii is known in the scientific community to be notoriously hard to work with, especially to genotype. This is due to the inability to establish a long-term *in vitro* cultivation system, with only one report of successfully culturing it [1]. Furthermore, because *Pneumocystis* species have a strict specificity for their respective host species, animal models for these fungi are hampered, significantly impacting research [2]. Therefore, molecular studies are largely dependent on DNA extracted directly from patient biospecimens [3], and the quality of the DNA obtained is influencingthe ability to detect *Pneumocystis* pneumonia (PCP), identify *P. jirovecii*, detect *P. jirovecii* colonisation and successfully genotype the causative isolates.

Pneumocystis can become an opportunistic pathogen when it infects immunocompromised hosts, but much is still unknown about *Pneumocystis* colonisation cases in both healthy and immunocompromised hosts, and their role in transmission and prevalence [2]. Confirming positive colonised cases is significantly more difficult than diagnosing PCP cases, due to the low fungal burden when compared with patients with PCP [4]. The gold standard for diagnosis for *Pneumocystis* infections for all species remains immunofluorescence staining from bronchoscopies or bronchoalveolar lavage (BAL), such as methenamine silver, Wright, or Giemsa, or the use of immunofluorescence-targeted antibodies, but these methods are only successful for diagnosis of PCP when the fungal load is high, such as in HIV positive patients [5]. Diagnosis employing conventional methods carry lower success rates when compared to molecular methods (e.g., PCR). Molecular methods are especially more effective in the detection of minute amounts of *P. jirovecii*, with successful rates largely seen using nested-PCR when paired with qPCR. They have been ableto predict better fungal burden prevalence in various geographical regions, in various age groupsand patient groups [5, 6].

Despite the advances made using molecular techniques, there are still gaps of knowledge in the epidemiology and pathogenesis of *Pneumocystis*, and this is especially true when trying to

determine the true prevalence of colonisation, the role of colonisation in the development of PCP infections, and the role of colonisation as a co-infection with comorbidities. Studies show significant variations when reporting colonisation numbers, ranging from 2.5% in Italy to 27.1% in Spain [3]. The range in the colonization rates could be due to several factors: (1) rates of underlyingmedical conditions in the community, (2) climate conditions in the different geographical regions,or (3) the molecular methods and technologies employed when undertaking identification andgenotyping [3, 7, 8].

2.1.1 Respiratory Specimens Used in *Pneumocystis* Diagnosis and Research

It is vital that the understanding behind *Pneumocystis* transmission and epidemiology is strengthened, especially regarding colonisation, as PCP carries up to 40% mortality inimmunocompromised hosts and identifying further risk factors may help prevent future infections. Various studies report unsuccessful sequencing analysis from positive PCP samples, with much discussion centred around the most adequate loci to be used for detection, screening, or genotyping,but little has been published about how to optimise *Pneumocystis* research from the very start – the specimen source sites and DNA extraction methodology. In 2019, Ruiz-Ruiz *et al.* published apaper discussing the effect of DNA extraction methods on *Pneumocystis* detection [2]. Althoughthis paper was highly informative, it also fell short, as it only examined two extraction methods, whilst working solely on DNA from lung tissue. Furthermore, papers which discuss the use of different biospecimens for DNA extraction, heavily focus on the specimen's reliability for identification and detection, not for their use in genotyping [9]. As such, it is important to discuss these variables, as their impact on successful *Pneumocystis* analysis is largely underestimated.

The source site of the *Pneumocystis* specimen can vary greatly, but also impact the amount of DNA which will be extracted. The *Pneumocystis* literature, which investigates detection and genotyping, heavily rely on lower respiratory specimens (e.g., lung tissue, bronchoalveolar washing

(BAL) or induced sputum (IS) samples), due to the fungal burden present in these specimens. The *Pneumocystis* trophozoites inhabit and replicate in the pulmonary alveoli, which results in lung tissue, especially alveolar tissue, producing the largest yield of fungal DNA. Despite the large quantities of DNA available from lung tissue, this procedure requires biopsies to obtain the sample, which is also the most invasive procedure, and therefore limits its availability. The downside of the Ruiz-Ruiz *et al.* study is, that it does not consider how best to extract DNA from more readily available biospecimens, such as BAL and IS specimens. BAL specimens currently remain the gold standard as *Pneumocystis* specimens for diagnosis and genotyping, increasing the necessity to identify pitfalls in extraction methods, which may be hindering genotyping [9].

BAL samples are performed via bronchoscopy using a saline solution to wash the pulmonary airways and the fluid is then captured for later use [10]. Even this remains an invasive procedure, BAL samples are more readily available than lung tissue, especially when done in hospitals for diagnosis of pulmonary infections. Additionally, BAL samples are more sensitive to detect *Pneumocystis* colonisers in non-symptomatic individuals [6]. Despite this, many patients will not willingly volunteer BAL specimens when there is no necessity, meaning that less invasive methods need also to be considered. IS samples are another type of biospecimen, which is often used in *Pneumocystis* research, as they are routinely used for diagnostics. The techniques to obtain IS specimens are less invasive than a bronchoscopy, as such these specimens are often the most accessible ones to researchers. To obtain a sputum sample, a patient inhales a nebulised sterile saline solution, which is then followed by expectoration for lower pulmonary secretions. A high-quality specimen will contain little saliva and should not be heavily contaminated by the host oral microbes. On the downside, there has been variability reported in the sensitivity of IS samples [11].

The main hurdle to all three techniques is their invasive nature, and that once DNA is extracted, it is always limited, and difficult to obtain further DNA from patients, as many are unwilling to have the procedure undertaken again. More recent literature has further explored the

options of using less invasive techniques, including the use of upper respiratory specimens, such as oropharyngeal washes (OPW), nasopharyngeal aspirate/washes (NPA), and nasopharyngeal swabs (NPS) [12-15].

Upper respiratory specimens are not recommended for diagnosis of PCP when using traditional microscopic techniques, due to the lack of sensitivity from lower fungal burden, but molecular methods have been shown to overcome this issue, especially when paired with nested PCR and qPCR [13]. The use of upper respiratory specimens has increased in recent times, largely due to their non-invasive nature and as they are less reliant on specialised training and equipment. The use of these specimens allows for (1) multiple samples to be collected at various time points,
(2) a quicker diagnosis, (3) less need for specialised training and equipment, and (4) for the surveillance of *P. jirovecii* colonisation in larger populations, without the implications of numerous or unnecessary invasive procedures. Promising, yet still fairly novel, more data on the use of these specimens needs to be acquired to strengthen their reliability, as information is still scant with slight inconsistencies between studies. Goterris *et al.* [16] reported an accuracy of 90% in PCP detection when using OPW samples and a sensitivity of 100%, whilst Juliano *et al.* [9] found lower rates, of 80% sensitivity and less than 70% specificity. This disparity is seen in multiple studies, with reports ranging as low as 50% sensitivity while others consistently report high levels well over 80%; with the specificity being recorded to range from 69% to 100% [17-20]. Hviid *et al.* also found that OPW samples had a high sensitivity, finding fungal carriage numbers in similarly to those obtained when using BAL samples, but found that in PCP patients, the use of OPW were significantly impaired if the patient was on prophylaxis or had even received one dose of treatment [12]. Goterris*et al.* [16] additionally suggested not using OPW specimens for colonisation detection in individuals with lower fungal load, which is common.

When using NPA specimens' high specificity and sensitivity levels have been reported across multiple studies, they also carry encouraging perspectives to be used in diagnosis, but like OPW

investigations also have varying reports in the literature. Kevin *et al.* [14] found 100% specificity and sensitivity, as well as both high negative and positive predictive values for PCP diagnosis. Samuel *et al.* [13] suggested PCR using NPA specimens should replace the gold standard techniques for diagnosis of PCP, which was also reinforced by Kelvin *et al.* [14], especially due to the simple procedure associated with the collection, as they are often collected at hospital admission for respiratory infections. This would also allow for a quick diagnosis, early treatment and stopping any possible nosocomial spread. Conversely, other studies showing lower specificity and sensitivityrates, such as that of Guigue *et al.* [15], contradict this view and warn that a negative NPAspecimen could not be used to exclude PCP, and should always be followed up with additional testing when PCP is suspected.

Scientists are increasingly overcoming the pitfalls of using upper respiratory specimens and advancements could soon lead to the regular inclusion of *P. jirovecii* in the panel of respiratory pathogens considered in hospitals. Detection of colonisation using these specimens is still debated, as many claim the C_T values assigned can be biased, blurring the lines between disease and colonisation, while struggling to detect fungal DNA in low level colonised individuals.

Despite the advancements made, these techniques cannot yet be reliably used for in-depth genotyping, and almost all literature explores using upper respiratory specimens focussing exclusively on the diagnostics in conjunction with qPCR and applying cut-off C_T values. Juliano *et al.* [9] successfully used OPW for microsatellite genotyping, but BAL and IS samples remain the most common samples when requiring DNA sequence analysis. By optimizing fungal DNA extraction methods for BAL and IS specimens, they could be translated across upper respiratory specimens, laying the groundwork for their potential use in genotyping analysis.

2.1.2 Improving Extraction of Fungal DNA and Volume of Lower Respiratory *Pneumocystis* Specimens

There are many different factors which may affect both the fungal load and DNA quality of *P. jirovecii* samples, which do not involve the source site or collection method. The largest influencer is the medical history of the patients - co-morbidities and the cause of their immunosuppression. Due to their low CD4+ cell count, patients which are HIV-positive have noticeably higher fungal loads, and therefore will produce specimens with a larger volume of DNA, making diagnosis easier than in patients with a PCP infection due to organ transplantation or cancer [4].

Another factor which has been commonly suspected to influence fungal load is the timepoint of the specimen collection in relation to prophylaxis and antifungal treatment. Reports have stated that as little as 24 hours after medical intervention the fungal DNA extracted from specimens could be radically affected, causing a false - negative result [16]. However, these factors are often listed as limitation, and described in detail, with advice on how to overcome these hurdles both in a clinical and research setting. Factors which are less researched are: (1) how molecular extraction methods impacting on acquiring fungal DNA, and (2) how to best optimise these techniques. Individual papers review individual methods, often in a specific relation to a commercially available kit, but there have been neither comparative nor optimising studies published, besides the Ruiz-Ruiz *et al.* study [2], which compared two techniques having been used on lung tissue samples.

There is an overwhelming amount of research undertaken using both upper and lower respiratory samples for diagnosis in conjunction with qPCR, from a variety of samples. Using qPCR allows for real time diagnosis but, naturally, does not supply genotyping information. Hviid *et al.* [12] in depth examine appropriate cut of C_T values, for a variety of upper respiratory samples, whilst also considering underlying disease, radiological signs (obtained by X-ray analysis and computed tomography scan), biological blood data (such as lymphocyte cell count and CD4/CD8 ratio), HIV burden (if applicable), results of direct physical examination), treatments, and clinical

outcomes. Guigue *et al.* [15] suggested that different C_T values should be assigned for upper and lower respiratory specimens, theorising that a high DNA value in specimens, such as NPA's, could be reflective of a higher fungal burden in the lungs than what has been reported previously.

When using fungal DNA for multilocus sequence typing (MLST) based genotyping it is important to obtain a large volume of high-quality DNA as possible. This is vital, as samples which produce low quality DNA have a greater chance of unreadable sequences and complement pairing, as well as an increasing risk of incorrect nucleotide assignment, which may impact subsequently the validity of the allele and sequence types assigned to the isolates. Specific allele types across genetic loci investigated in *Pneumocystis* MLST analysis are strongly correlated with certain patient characteristics, as seen with some cytochrome b (*CYB)* allele types being only recorded in patients with lung cancers [3], while others are significantly correlated with those undergoing prophylaxis [21]. Similarly, specific dihydropteroate synthase (*DHPS*) allele types are associated with resistance and mutations, able to assist clinicians with managing treatment. Incorrectly assigning a sequence type to a specimen could impact handling of outbreaks, accurate epidemiology reporting and even patient care. Considering IS and BAL samples are being mostly used in diagnosis and genotyping in both clinical and research settings, a review of currently used kits and how to best optimise them will greatly benefit those investigating *Pneumocystis.*

The Sputum DNA Isolation Kit (Norgen BioTek Corp., Thorold, ON, Canada) and theQIAamp DNA Mini Kit (Qiagen, Hilden, Germany) are the most used kits for the DNA extraction from *P. jirovecii*. Additionally, the commercially available QIAamp DNA Blood Kit (Qiagen, Hilden, Germany) was also analysed for samples which were heavily contaminated with blood. These three kits use a lysis buffer and Proteinase K digestion as the basis for extraction. Another commercially available kit being assessed was the Zymo Research Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, CA, USA), which is advertised as being designed for DNA isolation from "tough-to-lyse fungi", and uses bead beating for cell lysis, instead of denaturants and

proteinases. Lastly, an extraction and host-depletion, enrichment kit was also assessed, the Ultra Deep Microbiome Prep10 Molzym Enrichment Kit (Molzym GmbH & Co., Bremen, Germany). This kit is advertised as depleting host DNA to remove non-target host DNA to extract DNA only from the infecting agents (e.g., bacteria and fungi). To date, there have no successful enrichment methods been published for *Pneumocystis* and developing a highly effective protocol will help to overcome many downsides currently facing *Pneumocystis* research.

There are currently no optimization recommendations for the kits explored, when being used for DNA extraction from this fungal pathogen. Optimisation could be attempted in two ways, (1) adding an additional cell lysis step or (2) a host-DNA depletion step. Most commercial kits use two mechanisms for DNA extraction, the first being chemical and enzyme-based lysis, while others use mechanical destruction, such as through bead-beating. The cell wall of *Pneumocystis* has been characterised by Skalski *et al.* as “dynamic carbohydrate-rich structure consisting of β-glucans, chitins and other carbohydrate polymers” [22], and the thickness reaching its peak in the cyst form and is the thinnest whilst in the trophic form. The cell wall remains, like in other fungi, thick and harder to break through than cell walls of most prokaryotes. As the commercial kits being assessed were primarily using chemical and enzyme-based lysis, the extraction kits were also assessed for their efficiency when paired with a mechanical destruction step prior to applying the kit’s manufacturers protocol. A liquid nitrogen grinding step was included, before all the extraction kits have been used to both IS and BAL samples, and then the efficacy of the kit was compared to the results obtained when only using the manufacturer’s instructions. The second optimisation included a human DNA depletion step before the extraction using the commercial kits, by implementing the Saponin Host Depletion Protocol by Charalampous *et al.* [23].

Due to the low number of active PCP infections in Australia, specifically the greater Sydney, NSW region, it was difficult to obtain large quantities of fresh specimens, multiple, time points, and from different source sites for all patients. As such, the optimisation of the different kits was carried

out in-house on IS and BAL specimens, of both samples which had low positive *P. jirovecii* reads and those from confirmed PCP infections.

Furthermore, detection of PCP is also common in the veterinary health, often used for the detection of PCP in canines, especially seen in popular breeds, such as dachshunds and Cavalier King Charles Spaniels. As such, the efficiency of the kits and optimisation protocols when used to extract other *Pneumocystis* species was also assessed.

2.2 Methods

2.2.1 Samples and Extraction Method Evaluation

Samples were obtained from two hospitals in the greater Sydney region of NSW, Australia, Westmead Hospital and Royal Prince Alfred Hospital, between 2016 and 2017. A detailed list of the samples is shown in Table 2.1. Samples were sent to the Medical Mycology Research Laboratory for genotyping. The specimens were not from active infections of patients. Samples identified as "low fungal counts" were identified as *P. jirovecii* at the two hospitals diagnostic laboratories, using the QIAamp DNA Mini Kit from Qiagen as instructed, with no modifications, and qPCR targeting the *mt26S* genetic locus. No further patient information was given by the hospital, except the comorbidities (Table 2.1). A volume of 300 µl was taken from each sample, for testing each of the above-described kits, and each test condition.

Table 2.1. Human isolates which have been used in the evaluation and optimisation.

Patient ID	Source	PCP	Comorbidities
1022	BAL	+ low fungal count	Hep B+/cirrhosis
0993	BAL	Confirmed PCP	Kidney rejection and CMV
1459	BAL	Confirmed PCP	Liver transplant
1390	BAL	Confirmed PCP	Liver transplant
1952	IS	Confirmed PCP	Liver transplant
7334	IS	+ low fungal count	Lupus nephritis
1390	IS	Confirmed PCP	Liver transplant
4234	IS	Confirmed PCP	Liver transplant
1459	IS	Confirmed PCP	Liver transplant

Three BAL veterinary samples obtained from dogs with PCP, two from Sydney, NSW, and one from Melbourne, VIC, were also used for the evaluation of the different kits. Sample 1 and 2 from Sydney were taken by bronchoscopy from live dogs and stored at -80°C for 12 months; while Sample 3, from Melbourne, was a BAL specimen taken at an autopsy and delivered to the laboratory within 72 hours. The samples were contaminated with blood, especially the sample 2 from Sydney.

The twelve samples were used to evaluate five commercially available DNA extraction kits to extract *Pneumocystis* DNA. Kit 1: The Sputum DNA Isolation Kit, Kit 2: Qiagen - QIAamp DNA Mini Kit, Kit 3: Qiagen - QIAamp DNA Blood Kit, Kit 4: Zymo Research - Quick-DNA Fungal/Bacterial Miniprep Kit, and Kit 5: Molzym - Ultra Deep Microbiome PrepEnrichment Kit.

The following conditions were studied:

Condition 1: 300 µl of specimen were placed into 1.5 ml sterile Eppendorf tubes, and then the commercial kits were used according to the manufacturer's instructions and the DNA was eluted in 50 µl elution buffer.

Condition 2: 300 µl of specimen were placed into 1.5 ml sterile Eppendorf tubes. With the lid open, in a fume hood, a sterile pestle was placed inside the tube, and the tube was submerged 75% in liquid nitrogen for five seconds. The pestle was used to grind up and down inside the tube, rotating clockwise for thirty seconds, and then anti-clockwise for another thirty seconds. This was repeated for 3 minutes. After three minutes, the tube was submerged in liquid nitrogen again for three seconds, and the grinding procedure repeated for a following three minutes. The commercial kits were then used as per manufacturer's instructions and DNA was eluted in 50 µl elution buffer.

Condition 3: Host DNA depletion using saponin methodology taken from Charalampous *et al.,* 2019. 400 µl of the specimens were placed into 1.5 ml sterile Eppendorf tubes, and were centrifuged at 8,000 g for 5 min. Most of the supernatant were carefully removed (leaving up to50 µl of supernatant) and then remaining pellet was resuspended in 250 µl of PBS. Saponin

(Sigma-Aldrich, USA) was added to a final concentration of 2.5%, then mixed well and incubated at room temperature for 10 minutes. 350 µl of water was added the tube was incubated at room temperature for 30 seconds. After 12 µl of 5 M NaCl was added, to create an osmotic shock and lyse the host cells. Tubes were then again centrifuged at 6,000 g for 5 minutes. The supernatant is discarded and resuspended the pellet in 100 µl of phosphate-buffered saline (PBS). 100 µl of HL-SAN buffer (5.5 M NaCl and 100 mM $MgCl_2$ in nuclease-free water) was added, followed by 10 µl of HL-SAN DNase (ArcticZymes Technologies ASA, Norway). The specimens were incubated for 15 minutes at 37°C, with shaking at 800 revolutions per minute. This was done for host DNA digestion. The specimens were washed with 800 µl of PBS and centrifuged at 6,000 g for 3 minutes. This was repeated once more. The supernatant was then discarded, and the pellet resuspended in lysis buffer, and ready to be used for DNA extraction.

DNA concentrations were measured using DeNovix dsDNA Broad Range Fluorescent Assay, on the DS-11 Series Fluorometer (DeNovix Wilmington, DE, USA). DNA purity was also determined by measuring the ratios of absorbance at 260/280 and 280/230, to check if DNA extracts are contaminated by proteins and other organic particles, which may affect downstream procedures [2]. Statistical relevance was examined by performing t-tests and Mann-Whitney U tests on the datasets.

2.2.2 PCR Amplification

For the identification of *P. jirovecii* the genetic locus *mt26S* was amplified by PCR using the forward ‘5-TCAGGTCGAACTGGTGTACG-3’ and reverse primer ‘5- TGTTCCAAGCCCACTTCTT-3’. Samples were amplified in volumes of 25 µl per PCR reaction, using 10 X buffer (100 mM Tris-HCL, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$ and 0.01% w/v gelatine), 50 nM $MgCl_2$, 2 mM dNTPs, 10 ng/µl of each primer, 5 U/µl BIOTAQ DNA (Bioline) polymerase and 10 µl of genomic DNA. Amplification conditions were 95°C for 3 minutes for the

initial denaturing step; followed by 45 cycles of 94°C for 30 seconds, 55°C for 45 seconds and 72°C for 45 seconds; and finalised with 72°C for 7 minutes. The size of the amplified DNA bands was determined by agarose gel electrophoresis (1.5%) for 1 hour at 90 volts. Negative controls were used for each DNA extraction kit, and condition, detecting any possible contamination present in kit reagents.

2.3 Results

2.3.1 DNA Concentration and Extraction

The DNA of the twelve samples, when extracted using the five different kits, under 3 separate conditions was recorded in μg/ml, as well as their respective A260/280 and A260/230 ratios. Tables 2.2 to 2.6 show the individual DNA yield and purity ratios for each sample and used kit.

Table 2.2. Results from Kit 1: The Sputum DNA Isolation Kit**.** DNA extractions carried out under three different conditions, tested on 12 samples. C1 DNA extraction was carried out as per manufactures instruction. C2 DNA extraction was carried out with a liquid nitrogen grinding step before following manufacturers protocol. C3 DNA extraction was carries out with a saponin host-depletion step before following manufacturers protocol. Samples with DNA concentrations lower than 0.2 are recorded as <0.2.

Samples	**C1 Conc. µg/ml**	**C1 A260/ A280**	**C1 A260/ A230**	**C2 Conc. µg/ml**	**C2 A260/ A280**	**C2 A260/ A230**	**C3 Conc. µg/ml**	**C3 A260/ A280**	**C3 A260/ A230**
1022 - BAL	3.4	0.3	<0.2	30.2	1.11	1.56	<1.0	<0.2	<0.2
0993 - BAL	16.5	0.33	<0.2	36.8	1.24	1.32	<1.0	<0.2	<0.2
1459 - BAL	11.6	0.72	0.45	35.2	1.27	1.49	<1.0	<0.2	<0.2
1390 - BAL	8.9	0.89	0.6	40.3	1.69	1.01	<1.0	<0.2	<0.2
1952 - IS	35.6	1.06	0.84	105.4	1.8	1.93	<1.0	<0.2	<0.2
7334 - IS	21.6	0.81	0.79	90.3	1.62	2.09	<1.0	<0.2	<0.2
1390 - IS	29.8	1.02	0.9	98.7	1.67	1.89	<1.0	<0.2	<0.2
4234 - IS	34.7	0.99	0.92	102.3	1.75	1.78	<1.0	<0.2	<0.2
1459 - IS	29.9	0.78	0.68	100.1	1.68	1.87	<1.0	<0.2	<0.2
Dog 1 - BAL	26.4	0.46	0.5	46.9	1.32	1.55	<1.0	<0.2	<0.2
Dog 2 - BAL	11.1	0.58	0.53	35.2	1.19	1.49	<1.0	<0.2	<0.2
Dog 3 - BAL	40.5	1.02	0.64	111.6	1.43	1.36	<1.0	<0.2	<0.2

Table 2.3. Results from Kit 2: QIAamp DNA Mini Kit. DNA extractions carried out under three different conditions, tested on 12 samples. C1 DNA extraction was carried out as per manufactures instruction. C2 DNA extraction was carried out with a liquid nitrogen grinding step before following manufacturers protocol. C3 DNA extraction wascarries out with a saponin host-depletion step before following manufacturers protocol. Samples with DNA concentrations lower than 0.2 are recorded as <0.2.

Samples	**C1 Conc. µg/ml**	**C1 A260/ A280**	**C1 A260/ A230**	**C2 Conc. µg/ml**	**C2 A260/ A280**	**C2 A260/ A230**	**C3 Conc. µg/ml**	**C3 A260/ A280**	**C3 A260/ A230**
1022 - BAL	35.9	1.32	1.57	99.7	1.65	1.48	<1.0	<0.2	<0.2
0993 - BAL	41.2	0.99	1.23	132.2	1.73	1.83	<1.0	<0.2	<0.2
1459 - BAL	27.3	1.19	1.3	105.5	1.94	1.94	<1.0	<0.2	<0.2
1390 - BAL	29.5	0.93	0.89	90.3	1.9	1.33	<1.0	<0.2	<0.2
1952 - IS	35.6	1.03	1.1	122.2	1.84	1.96	<1.0	<0.2	<0.2
7334 - IS	37.9	1.21	1.05	117.4	1.79	1.89	<1.0	<0.2	<0.2
1390 - IS	29.8	0.98	1.01	80.3	1.77	1.66	<1.0	<0.2	<0.2
4234 - IS	35.0	1.17	0.99	113.4	2.03	1.77	<1.0	<0.2	<0.2
1459 - IS	31.7	0.98	1.0	154.3	1.64	1.71	<1.0	<0.2	<0.2
Dog 1 - BAL	41.6	0.99	1.09	88.9	1.68	1.42	<1.0	<0.2	<0.2
Dog 2 - BAL	42.3	1.32	1.12	97.3	1.33	1.49	<1.0	<0.2	<0.2
Dog 3 - BAL	89.3	1.38	1.49	204.4	1.72	1.99	<1.0	<0.2	<0.2

Table 2.4. Results from Kit 3: QIAamp DNA Blood Kit**.** DNA extractions carried out under three different conditions, tested on 12 samples. C1 DNA extraction was carried out as per manufactures instruction. C2 DNA extraction was carried out with a liquid nitrogen grinding step before following manufacturers protocol. C3 DNA extraction wascarries out with a saponin host-depletion step before following manufacturers protocol. Samples with DNA concentrations lower than 0.2 are recorded as <0.2.

Samples	**C1 Conc. µg/ml**	**C1 A260/ A280**	**C1 A260/ A230**	**C2 Conc. µg/ml**	**C2 A260/ A280**	**C2 A260/ A230**	**C3 Conc. µg/ml**	**C3 A260/ A280**	**C3 A260/ A230**
1022 - BAL	4.4	0.4	0.38	42.1	1.44	1.5	<1.0	<0.2	<0.2
0993 - BAL	5.7	0.23	<0.2	12.4	1.36	1.02	<1.0	<0.2	<0.2
1459 - BAL	7.7	0.53	0.6	38.2	0.8	0.45	<1.0	<0.2	<0.2
1390 - BAL	5.4	0.6	0.42	44.9	1.66	1.28	<1.0	<0.2	<0.2
1952 - IS	<1.0	<0.2	<0.2	3.3	0.25	0.3	<1.0	<0.2	<0.2
7334 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	0.28	<1.0	<0.2	<0.2
1390 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
4234 - IS	<1.0	<0.2	<0.2	2.9	3.6	0.33	<1.0	<0.2	<0.2
1459 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
Dog 1 - BAL	2.3	<0.2	<0.2	55.3	1.66	1.72	<1.0	<0.2	<0.2
Dog 2 - BAL	33.4	1.34	1.1	86.3	1.69	1.92	<1.0	<0.2	<0.2
Dog 3 - BAL	11.8	0.9	0.98	104.3	1.37	1.22	<1.0	<0.2	<0.2

Table 2.5. Results from Kit 4: Zymo Research Quick-DNA Fungal/Bacterial Miniprep**.** DNA extractions carried out under three different conditions, tested on 12 samples. C1 DNA extraction was carried out as per manufactures instruction. C2 DNA extraction was carried out with a liquid nitrogen grinding step before following manufacturers protocol. C3 DNA extraction was carries out with a saponin host-depletion step before following manufacturersprotocol. Samples with DNA concentrations lower than 0.2 are recorded as <0.2.

Samples	**C1 Conc. µg/ml**	**C1 A260/ A280**	**C1 A260/ A230**	**C2 Conc. µg/ml**	**C2 A260/ A280**	**C2 A260/ A230**	**C3 Conc. µg/ml**	**C3 A260/ A280**	**C3 A260/ A230**
1022 - BAL	12.8	1.01	0.98	28.3	1.32	1.5	<1.0	<0.2	<0.2
0993 - BAL	15.9	0.68	1.03	30.4	1.58	1.43	<1.0	<0.2	<0.2
1459 - BAL	11	1.17	0.96	24.8	1.43	1.52	<1.0	<0.2	<0.2
1390 - BAL	27.4	1.15	1.07	38.5	1.34	1.64	<1.0	<0.2	<0.2
1952 - IS	2.1	0.52	0.49	8.9	0.98	0.88	<1.0	<0.2	<0.2
7334 - IS	3.8	0.44	0.29	6.1	0.73	0.68	<1.0	<0.2	<0.2
1390 - IS	3.4	0.63	0.44	11.2	0.99	0.31	<1.0	<0.2	<0.2
4234 - IS	2.9	0.58	0.63	8.9	0.64	0.44	<1.0	<0.2	<0.2
1459 - IS	4.8	0.72	0.7	16.7	0.82	0.91	<1.0	<0.2	<0.2
Dog 1 - BAL	12.8	1.18	1.5	26.8	1.33	1.48	<1.0	<0.2	<0.2
Dog 2 - BAL	8.2	1.23	1.36	27.7	1.49	1.02	<1.0	<0.2	<0.2
Dog 3 - BAL	45.2	1.41	1.3	60.8	1.52	1.38	<1.0	<0.2	<0.2

Table 2.6. Results from Kit 5: Ultra Deep Microbiome Prep10 Molzym Enrichment Kit**.** DNA extractions carried out under three different conditions, tested on 12 samples. C1 DNA extraction was carried out as per manufactures instruction. C2 DNA extraction was carried out with a liquid nitrogen grinding step before following manufacturers protocol. C3 DNA extraction was carries out with a saponin host-depletion step before following manufacturersprotocol. Samples with DNA concentrations lower than 0.2 are recorded as <0.2.

Samples	**C1 Conc. µg/ml**	**C1 A260/ A280**	**C1 A260/ A230**	**C2 Conc. µg/ml**	**C2 A260/ A280**	**C2 A260/ A230**	**C3 Conc. µg/ml**	**C3 A260/ A280**	**C3 A260/ A230**
1022 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
0993 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
1459 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
1390 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
1952 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
7334 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
1390 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
4234 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
1459 - IS	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
Dog 1 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
Dog 2 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2
Dog 3 - BAL	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2	<1.0	<0.2	<0.2

Due to the lack of successful DNA extraction achieved when using Kit 5, as well as when using all the kits with the inclusion of the saponin host-depletion protocol (C3), the results for Kit 5 and C3 have not been considered for further analysis.

Figure 2.1 visualises the DNA concentration levels for kits 1 to 4, under conditions 1 and 2, for each sample, showing that Kit 2: QIAamp DNA Mini Kit, when combined with the liquid nitrogen grinding step, yielded the highest DNA concentration.

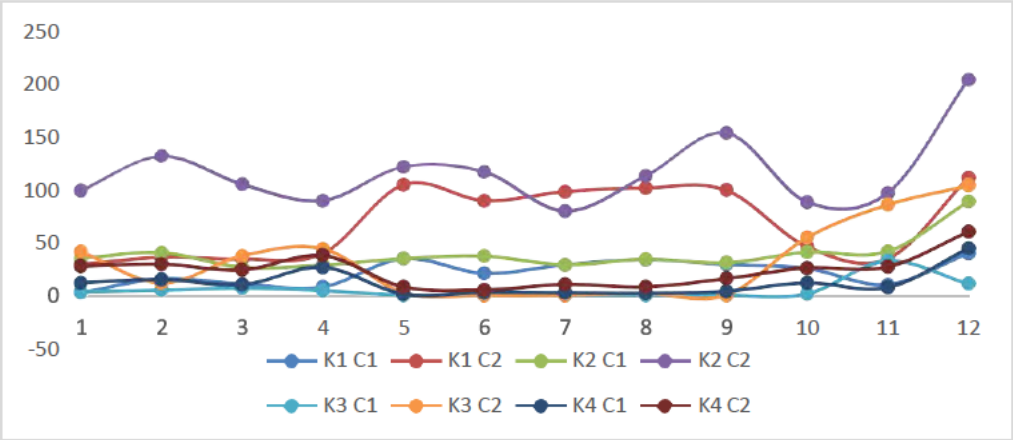


Figure 2.1. The DNA concentration (µg/ml) extracted from kit 1 – 4, when using the manufactures protocol (C1), and with an inclusion of the liquid nitrogen grinding step before following the manufactures protocol (C2).

The mean scores of the twelve samples were calculated, seen in Table 2.7 and Figure 2.2. The average numbers reinforce that the QIAamp DNA Mini Kit with the mechanical grinding step, also produced the purest DNA quality, obtaining scores closest to 1.8.

Table 2.7. Mean scores of kits, under respective conditions, excluding Kit 5.

Mean	K1 C1	K1 C2	K2 C1	K2 C2	K3 C1	K3 C2	K4 C1	K4 C2
Concentration µg/ml	26.1	80.02	41.25	118.57	6.42	33.27	12.31	22.97
A260/A280	0.75	1.48	1.12	1.75	0.67	1.5	0.89	1.18
A260/A230	0.66	1.61	1.15	1.71	0.7	1	0.9	1.1

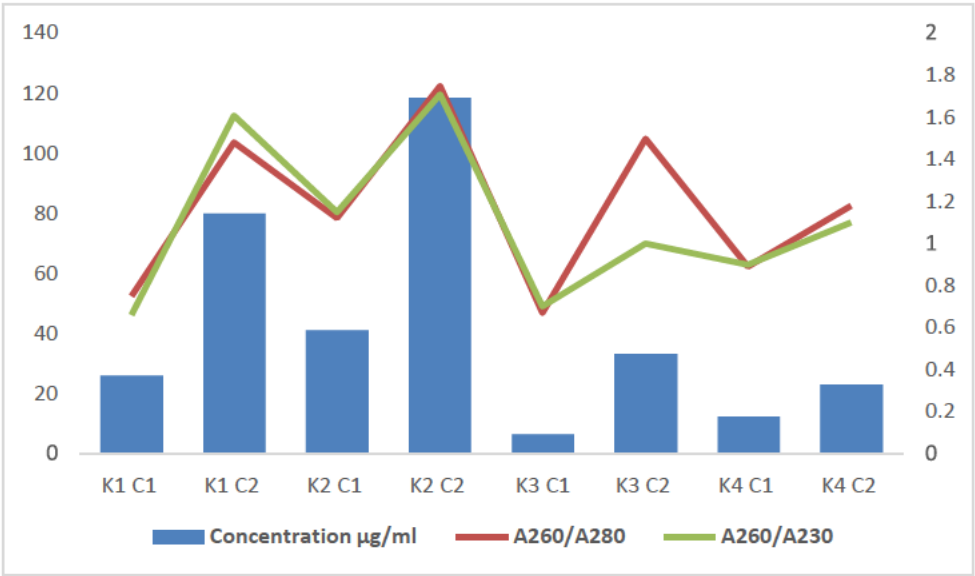


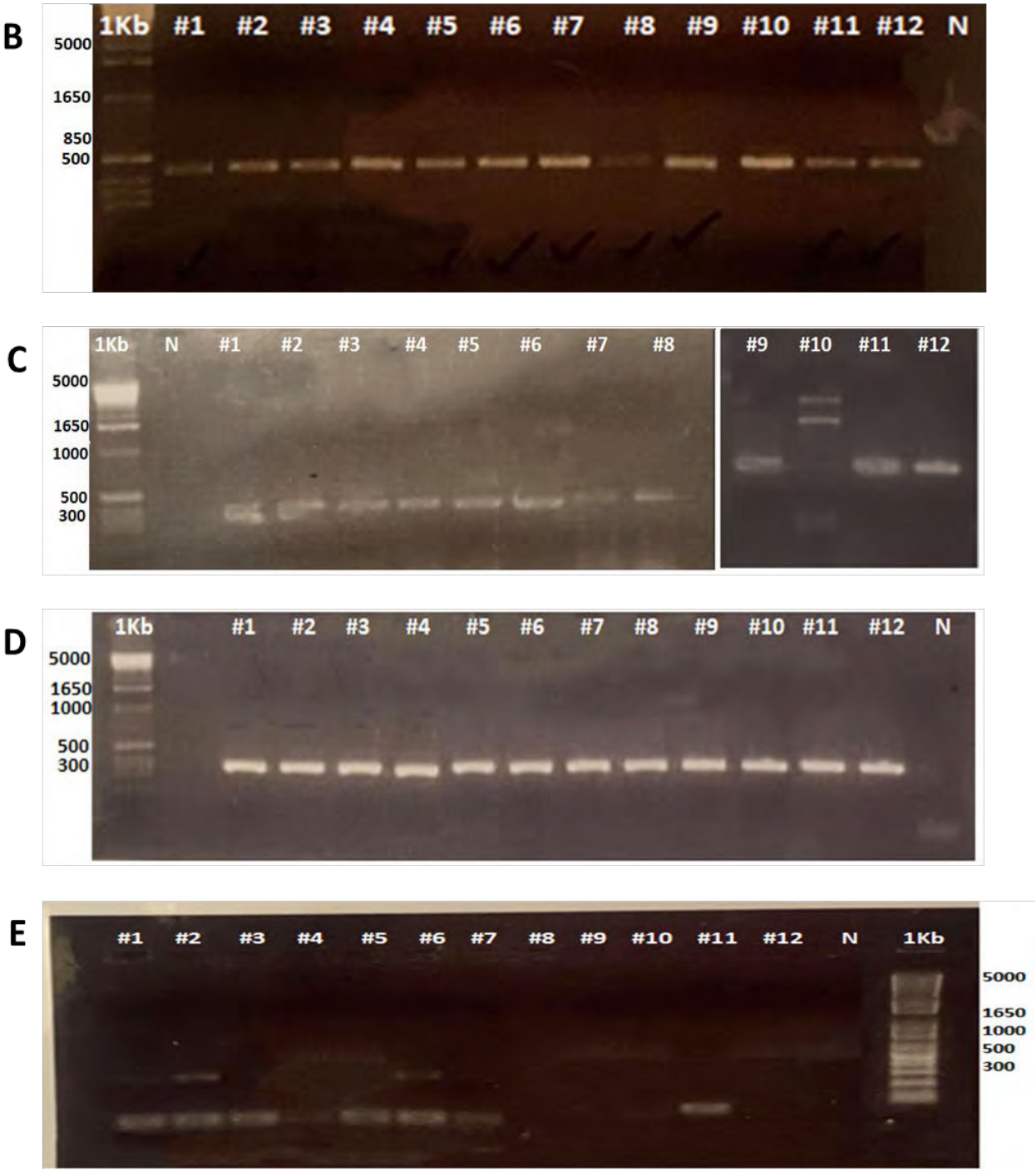
Figure 2.2. The mean DNA concentration from Kits 1 -4, under conditions 1 and 2 are shown as blue bars on the horizontal axis, with the nucleic acid purity scores displayed on the vertical axis.

When combined with the liquid nitrogen grinding step all four kits show a significant increase of DNA yield ($p < .00001$, with a Z score of -4.42222, and U value of 548) when compared with the DNA concentrations obtained using the kits as per manufacturers protocol only. Individually, Kits 1 demonstrated a statistical significance of $p = 0.0014$ between the two conditions, this was also the case for Kit 2 ($p < .00001$), for Kit 3 ($p = 0.0059$), while Kit 4 did not show a statistical significance ($p = 0.0571$) and yielded the lowest DNA concentration when combined with the liquid nitrogen protocol.

Kit 1: The Sputum DNA Isolation Kit performed overall secondary to Kit 2, but there was still a significant difference of the overall performance between the two kits, when combined with the liquid-nitrogen grinding step (C2) ($p = 0.0028$). The BAL specimens performed significantly worse than the IS samples, even when isolated from the same patient and using Kit 1 ($p = 0.0029$), whilst this difference between the two specimens was not noted using Kit 2 ($p = 0.9771$).

2.3.2 Gel Electrophoresis

The twelve samples, from Kits 1 – 4, with no deviation from the manufacturer's protocols, and then with the addition of the liquid nitrogen grinding to the protocol were visualised with gel electrophoresis, when amplifying the genetic locus *mt26S* (Figure 2.4).


B
1Kb #1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12 N
5000
1650
850
500
C
1Kb N #1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12
5000
1650
1000
500
300
D
1Kb #1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12 N
5000
1650
1000
500
300
E
#1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12 N 1Kb
5000
1650
1000
500
300

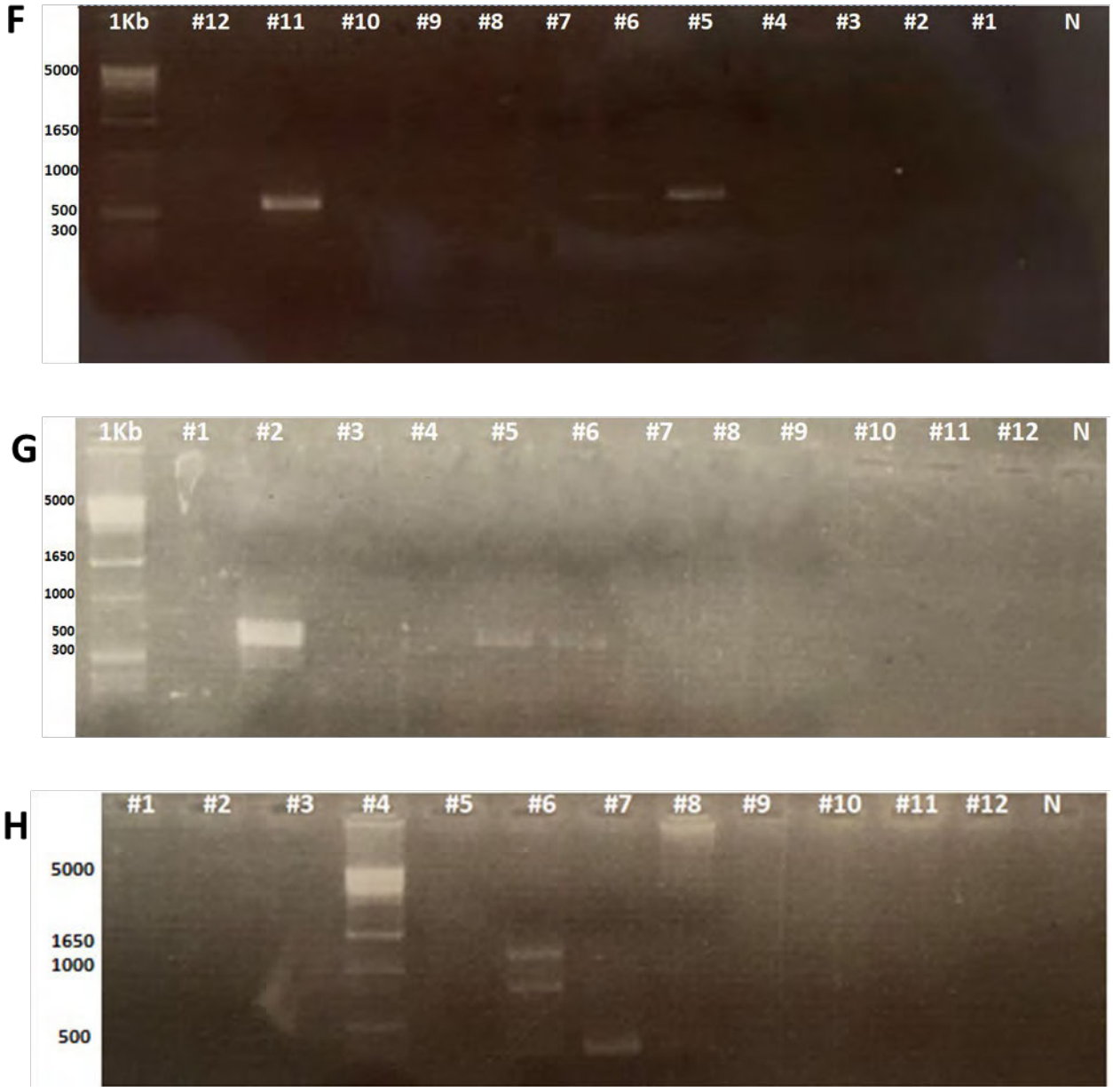


Figure 2.4. A) Kit 1, Condition 1 **B)** Kit 1, Condition 2 **C)** Kit 2, Condition 1 **D)** Kit 2, Condition 2 **E)** Kit 3, Condition 1 **F)** Kit 3, Condition 2 **G)** Kit 4, Condition 1 **H)** Kit 4, Condition 2**.** All images have been cropped and edited, but only to combine gel wells from separate gel images, and to add captions. Sample 1) 1022 – BAL, 2) 0993 – BAL, 3) 1459 – BAL, 4) 1390 – BAL, 5) 1952 – IS, 6) 7334 – IS 7) 1390 – IS, 8) 4234 – IS, 9) 1459 – IS, 10) Dog 1 – BAL, 11) Dog 2 – BAL, 12) Dog 3 -BAL. N is indicative of the negative control used. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen, USA) has been used and labelled where possible.

The strongest and thickest bands visualised showing an amplified PCR product, which was obtained after the DNA had been extracted using Kit 2, when combined with condition 2. All 12 samples show strong, thick bands at the correct fragment size of approximately 300 bp (Figure

2.4.D). When extracted as per manufactures protocol and then amplified, the bands were not as bright, with faintly smeared bands also visualised for certain samples (Figure 2.4.C).

Kit 1 under condition 2, had the strongest bands for the IS samples, with more faint bands visualised for the BAL samples, but the samples were of worse quality when extracted solely using the manufactures protocol, and then amplified, producing both multiple bands and smeared bands. Kit 3 and Kit 4 under both conditions (Figure 2.4.E-H), had the largest number of samples which were unsuccessfully amplified, such as PCR products which were faint, smeared or appearing in multiple bands. An outlier to this is sample 11 (Dog 2 – BAL), which was a BAL specimen heavily contaminated with blood. When extracted by Kit 3 processed under condition 2 (Figure 2.4.F), the PCR product showed a bright and clearly amplified band. The band is comparable to the amplification of sample 11 when extracted with Kit 2 and processed under condition 2 (Figure 2.4.D).

2.4 Discussion

Pneumocystis is notoriously a difficult fungus to work with, stemming from the inability to readily being cultured *in vitro*, while its strict species specificity makes it difficult to use *in vivo* models. The importance of optimising the DNA extracted from patient specimens is paramount, but often overlooked in research, and not very widely explored. A vital step when commencing any *P. jirovecii* investigations is extracting DNA from the various upper and lower respiratory tractsamples, which will in turn influence all downstream procedures and results. There have been numerous publications regarding the type of specimen used for PCP diagnosis, with contradictory results reported. Very seldom are the various specimens discussed in the context of DNA extraction methodology/protocol, and how to optimise the DNA concentration and quality, despite this being discussed and reported more frequently with other fungi [24-26], and often in the contexts of bacterial DNA extraction [27-29]. There are no guidelines on how to best extract DNA and how to

optimise the extraction protocols, except for the limited Ruiz-Ruiz *et al.* study for *P. jirovecii*. The Ruiz-Ruiz *et al.* study showed that an alteration to one commercial extraction kit produced a significant higher concentration of DNA and purity when extracted from lung biopsy specimens. As such, it was important to examine the effect on DNA extraction when different popular commercial kits were used on the most common respiratory samples used in *P. jirovecii* research – BAL and IS. The results from these experiments are presented as values of the DNA extraction concentration, and the purity of the DNA. The DNA was then used to amplify the genetic locus, the mitochondrial rRNA gene, *mt26S*, and visualised on gel electrophoresis, to confirm that the DNA extracted contained *P. jirovecii*. The genetic locus *mt26S* was selected as it is a multicopy gene with no reports of difficulty amplifying and is often used in PCP diagnosis. The obtained results showedthat the extraction methodology is significantly altering the concentration and purity of the isolates obtained, and in turn impacts their ability to be amplified by PCR, and potentially effects the obtained sequencing quality.

Twelve *P. jirovecii* specimens comprising of both IS and BAL samples were extracted using five different commercial kits in combination with three different conditions. The QIAamp DNA Mini Kit (Kit 2) when combined with a liquid nitrogen grinding step extracted significantly higher concentrations than any of the other kits, along with having the highest A260/A280 and A260/A230 purity scores. Additionally, also the strongest bands after visualization on gel electrophoresis were noted from this Kit using the condition 2, with all bands brightly amplified. Although the QIAamp manufacturers protocol states “No mechanical homogenization is necessary as the tissues are lysed enzymatically”, our results show that a mechanical grinding step completed before undertaking the extraction protocol, raises the yield of DNA significantly and impacts downstream applications. The Norgen Sputum DNA Isolation Kit along with the QIAamp DNA Blood Kit also achieved significantly higher concentration of DNA and purer DNA when including a mechanical grinding step. The Zymo Research Quick-DNA Fungal/Bacterial Miniprep Kit was the only kit which did not yield significantly higher levels of DNA concentration when including a liquid nitrogen

grinding step in the protocol and yielded the least DNA in this test series. As noted previously, the Zymo Research Kit was the only kit which used mechanical lysis, in the form of bead beating for the lysis of the cells and no chemical lysis, which was used by the other three extraction kits (not Kit 5). This indicates that fungi, such as *Pneumocystis,* require both chemical lysis alongside a mechanical lysis step for successful DNA extraction. Both the QIAamp DNA Mini and Norgen Kit used the combination of enzyme digestion in the form of Proteinase K and a lysis buffer for the chemical lysis, and both performed better than the Zymo Research Kit, when used without alterations. This strongly suggests that chemical lysis is more effective than mechanical lysis, in the form of bead beating, when used individually and when paired together.

Noticeably, when using the Sputum DNA Isolation kit, significantly higher levels of DNA concentration were extracted from IS specimens than from BAL samples, except for the Dog sample 3. The kit specialises in the extraction of DNA from sputum samples, containing “a sputum liquefier to reduce viscosity and facilitate DNA isolation” according to the manufacturers protocol, which is reflective in the results obtained. When comparing only the IS samples, there was no significant difference ($p = 0.1696$) between the samples when extracted using the Norgen Kit and the QIAamp Kit (when combined with liquid nitrogen grinding). Promisingly though, when paired with a mechanical grinding step, the Norgen Kit amplified the *mt26S* locus for all samples. Conversely, when only looking at the DNA extraction values, the QIAamp DNA Blood Kit was unable to successfully extract DNA from IS samples (as 3.3 μg/ml was the highest) but successfullyextracted DNA from BAL samples (with a maximum 104.3 μg/ml yield from the Dog Sample 3). The kit also yielded a high DNA concentration from the Dog sample 2, which was the only sample heavily contaminated by blood, with comparable levels with those from the QIAamp DNA MiniKit. Unfortunately, only the Dog sample 2 and sample 7334 (IS sample) successfully amplified a band when visualised on gel electrophoresis, and since human sample 7334 measured very low DNA concentration, it is possible that the amplified band is only contamination or human error. Thehigh levels of DNA concentration extracted from some samples, such as for the Dog sample 3,

could be indicative of a high DNA yield of non-*Pneumocystis* microbes. This kit did not produce a satisfactory amplification to warrant its use in *P. jirovecii* research or detection; but its use in *P. jirovecii* samples with a high blood contamination may be advisable after further testing and only if similar results are obtained as for the Dog sample 2.

There were also two patients which provided both IS and BAL samples, 1459 and 1390, whilst in hospital admission. Although it is often stated that BAL samples result in higher DNA yields, this statement has been questioned in more recent studies [16]. When observing the results of the two specimen types using the QIAamp DNA Mini Kit, sample 1495 yielded greater amounts of DNA from the IS sample, while the sample 1390 yielded greater amounts from the BAL sample. When DNA was extracted from the two BAL samples higher A260/280 and A260/230 ratios where obtained, indicating a greater DNA purity. Additionally, no distinction between bands was evident when looking at both Kit 1 and 2, under condition 2. Albeit a small sample size, there is no indication throughout the results that either specimen type results in a greater DNA concentration. The QIAamp DNA Mini Kit is the most universal of the four kits, advertised to work across a large variety of host specimen types, including respiratory and blood sample types, as well as being the least time-consuming of all kits (20 minutes DNA extraction *vs* 30 minutes DNA extraction fromthe Norgen Kit), while all others taking closer to an hour. Its applicability to a larger variety of samples, regardless of the host species, makes it the most accessible kit in a diverse laboratory or hospital setting.

Unfortunately, neither the enrichment or host DNA depletion protocols were successful with extracting and subsequently amplifying DNA from the human samples. The Ultra Deep Microbiome Prep10 Molzym Enrichment Kit obtained no DNA from both types of samples, despitethe kit being advertised to combine “several protocols” for work on “ultra-sensitive” microorganisms, and for use on BAL samples. It is possible that the depletion step removed all fungal DNA, due to the minimal amounts present, or that there was human error or kit malfunction.

Similarly, no extraction was possible when using the saponin host-depletion step (condition 3) before using the commercial kits, which previously were able to obtain DNA. Even the Saponin protocol has showed very promising results when used for bacterial detection from complex human clinical samples, it could not be successfully applied to fungal detection in similar samples, possiblydue to the fact that it also removed the fungal DNA, and therefore is not suitable for eukaryotic microorganism. Alternatively, it is important noting, that both the Molzym Kit and the saponin step were the last of the extraction protocols to be completed, with the specimens previously frozen and then rethrown several times when needed for extraction by the other kits and test conditions. It may be that the lack of DNA extraction is caused by DNA degradation from the temperature changes. Future research should be undertaken on enrichment protocols using fresh specimens to confidently rule out these problems.

It is interesting to note that the Dog sample 3 amplified considerably stronger bands from DNA obtained with all extraction methods tested than any of the other human or dog samples. Itwas extracted from a lung biopsy sample from a deceased dog, which contained multiple DNA heavy trophozoites. Interestingly, in instances when there were high DNA yields, the sample did notalways amplify the *mt26S* band. On the contrary, the two isolates which were originally flagged as “low level fungal counts”, sample 1022 and sample 7334, amplified successfully when usingQIAamp DNA Mini Kit with the added the liquid-nitrogen grinding step, and produced PCR bands comparable to all other specimens.

The data collected from these experiments strongly suggest that the sample source site is not a major influencer of the DNA concentration or purity of the sample, which was often theorised. In fact, the DNA extraction methods (e.g., commercial kits) play a much more crucial role in achieving a high concentration and quality of the DNA extracted.

Further factors hypothesized to impact DNA yield are the patient’s co-morbidities, PCP treatment and prophylaxis history, as well as transportation and storage of the respiratory sample.

Although the metadata for these patients were very limited, external factors had not be noted to significantly influence extraction and subsequent amplification of the *mt26S* locus when observing results obtained from DNA extracted with the QIAamp DNA Mini Kit, as all twelve samples successfully amplified a PCR product. A detailed patient history alongside more extensive analysis of the samples across multiple genetic loci, involving both qPCR and DNA sequencing of the isolates, would be necessary to further comment on the extent external factors play on the DNA quality extracted.

Drawbacks of this study are that it was a retrospective study, lacking fresh specimens, as well as working with a limited sample size. Future analysis would benefit from collecting as many lower and upper respiratory samples at the start of the patient's admission from a larger cohort, as well as to use qPCR and sequencing of all samples.

2.5 Conclusion

The combination of both chemical and mechanical lysis when using a versatile extraction kit, such as the QIAamp DNA Mini Kit, allows for significantly higher levels of DNA concentrations to be extracted, as well as increase the purity of the extracted DNA. The QIAamp DNA Mini Kit when combined with a liquid nitrogen grinding step successfully amplified the *mt26S* locus for all twelve samples, despite host species, source site or the presence of contaminants. It is advisable that all isolates queried for either identification or genotyping of *P. jirovecii* from lower respiratory samples utilize the QIAamp DNA Mini Kit with the addition of the mechanical homogenisationstep before.

2.6 References

1. Schildgen, V., *et al., Pneumocystis jirovecii can be productively cultured in differentiated CuFi-8 airway cells.* MBio, 2014. 5(3).

2. Ruiz-Ruiz, S., *et al., New DNA Extraction Method for the Detection of Pneumocystis in Lung Tissue Samples of Colonized Individuals.* 2019.

3. Sokulska, M., *et al., Genotyping of Pneumocystis jirovecii in colonized patients with various pulmonary diseases.* Medical mycology, 2018. 56(7): p. 809-815.

4. Fauchier, T., *et al., Detection of Pneumocystis jirovecii by quantitative PCR to differentiate colonization and pneumonia in immunocompromised HIV-positive and HIV-negative patients.* Journal of clinical microbiology, 2016. 54(6): p. 1487-1495.

5. Morris, A., *et al., Epidemiology and Clinical Significance of Pneumocystis Colonization.* The Journal of Infectious Diseases, 2008. 197(1): p. 10-17.

6. Botterel, F., *et al., Clinical significance of quantifying Pneumocystis jirovecii DNA by using real-time PCR in bronchoalveolar lavage fluid from immunocompromised patients.* Journal of clinical microbiology, 2012. 50(2): p. 227-231.

7. Varela, J., *et al., Climatic factors and Pneumocystis jiroveci infection in southern Spain.* Clinical microbiology and infection, 2004. 10(8): p. 770-772.

8. Ponce, C.A., *et al., Pneumocystis Colonization Is Highly Prevalent in the Autopsied Lungs of the General Population.* Clinical Infectious Diseases, 2010. 50(3): p. 347-353.

9. Juliano, J.J., *et al. Use of oropharyngeal washes to diagnose and genotype Pneumocystis jirovecii*. in *Open forum infectious diseases*. 2015. Oxford University Press.

10. Liebler, J.M. and C.J. Markin, *FIBEROPTIC BRONCHOSCOPY FOR DIAGNOSIS AND TREATMENT.* Critical Care Clinics, 2000. 16(1): p. 83-100.

11. Babb-Biernacki, S.J., J.A. Esselstyn, and V.P. Doyle, *Rethinking host range in Pneumocystis.* PLOS Pathogens, 2020. 16(9): p. e1008824.

12. Hviid, C.J., *et al., Detection of Pneumocystis jirovecii in oral wash from immunosuppressed patients as a diagnostic tool.* PloS one, 2017. 12(3): p. e0174012.

13. Samuel, C.M., *et al., Improved detection of Pneumocystis jirovecii in upper and lower respiratory tract specimens from children with suspected pneumocystis pneumonia using real-time PCR: a prospective study.* BMC infectious diseases, 2011. 11(1): p. 1-6.

14. Kelvin, K.W.T., *et al., Use of Nasopharyngeal Aspirate for Diagnosis of Pneumocystis Pneumonia.* Journal of Clinical Microbiology, 2013. 51(5): p. 1570-1574.

15. Guigue, N., *et al., Utility of adding Pneumocystis jirovecii DNA detection in nasopharyngeal aspirates in immunocompromised adult patients with febrile pneumonia.* Medical mycology, 2015. 53(3): p. 241-247.

16. Goterris, L., *et al., Molecular Diagnosis of Pneumocystis jirovecii Pneumonia by Use of Oral Wash Samples in Immunocompromised Patients: Usefulness and Importance of the DNA Target.* Journal of clinical microbiology, 2019. 57(12).

17. Wakefield, A., *et al., DNA amplification on induced sputum samples for diagnosis of Pneumocystis carinii pneumonia.* The Lancet, 1991. 337(8754): p. 1378-1379.

18. Helweg-Larsen, J., *et al., Diagnostic Use of PCR for Detection of Pneumocystis carinii in Oral Wash Samples.* Journal of clinical microbiology, 1998. 36(7): p. 2068-2072.

19. Matos, O., *et al., Effect of oral washes on the diagnosis of Pneumocystis carinii pneumonia with a low parasite burden and on detection of organisms in subclinical infections.* European Journal of Clinical Microbiology and Infectious Diseases, 2001. 20(8): p. 573- 575.

20. Atzori, C., *et al., Combined use of blood and oropharyngeal samples for noninvasive diagnosis of Pneumocystis carinii pneumonia using the polymerase chain reaction.* European Journal of Clinical Microbiology and Infectious Diseases, 1998. 17(4): p. 241- 246.

21. Szydłowicz, M., *et al., Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients—preliminary report.* Parasitology research, 2019. 118(1): p. 181-189.

22. Skalski, J.H., T.J. Kottom, and A.H. Limper, *Pathobiology of Pneumocystis pneumonia: life cycle, cell wall and cell signal transduction.* FEMS Yeast Research, 2015. 15(6).

23. Charalampous, T., *et al., Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection.* Nature biotechnology, 2019. 37(7): p. 783-792.

24. Muñoz-Cadavid, C., *et al., Improving Molecular Detection of Fungal DNA in Formalin-Fixed Paraffin-Embedded Tissues: Comparison of Five Tissue DNA Extraction Methods Using Panfungal PCR.* Journal of Clinical Microbiology, 2010. 48(6): p. 2147-2153.

25. Griffiths, L.J., *et al., Comparison of DNA extraction methods for Aspergillus fumigatus using real-time PCR.* Journal of Medical microbiology, 2006. 55(9): p. 1187-1191.

26. Metwally, L., *et al., Improving molecular detection of Candida DNA in whole blood: comparison of seven fungal DNA extraction protocols using real-time PCR.* Journal of medical microbiology, 2008. 57(3): p. 296-303.

27. McOrist, A.L., M. Jackson, and A.R. Bird, *A comparison of five methods for extraction of bacterial DNA from human faecal samples.* Journal of microbiological methods, 2002.50(2): p. 131-139.

28. Claassen, S., *et al., A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples.* Journal of microbiological methods, 2013. 94(2): p. 103-110.

29. Queipo-Ortuno, M., *et al., Comparison of seven commercial DNA extraction kits for the recovery of Brucella DNA from spiked human serum samples using real-time PCR.* European Journal of Clinical Microbiology & Infectious Diseases, 2008. 27(2): p. 109-114.

Chapter 3

Optimization of the Original ISHAM MLST Scheme and Proof of Concept Application of the Optimized MLST Scheme to a Set of Chilean Clinical Samples

3.1 Introduction

Epidemiological investigations of *Pneumocystis* have been less successful than epidemiological studies for other nosocomial fungal infections due to the inability to isolated, grow, and sustain *Pneumocystis* in pure culture. Over the past two decades, molecular studies and genotype investigations have helped to better understand the prevalence, transmission, to guide treatment regimens, understand drug resistance, virulence, cell biology and pathogenesis of *Pneumocystis* [1]. Genotyping of nosocomial outbreaks is crucial for understanding the epidemiology of *Pneumocystis* infections, and the continuous development of public health approaches towards preventing and curbing future outbreaks [2]. Genotyping has provided valuable information about the genetic diversity of *Pneumocystis*, illustrating differences between species and genotypes, clinical outcomes, geographic distribution, modes of transmission, and population genetics [3]. Multilocus sequence typing (MLST) has become the benchmark technique for genotyping, generating many various MLST schemes to be used to investigate *Pneumocyctis* Pneumonia (PCP) [4]. An outbreak of PCP within a renal transplant unit at Westmead Hospital, Sydney, NSW, Australia, in 2011 led tothe development of a novel MLST scheme for genotyping *P. jirovecii,* which then led to the development of the very first MLST database dedicated to *P. jirovecii* under the auspice of the International Society of Human and Animal Mycology (ISHAM,"http://mlst.mycologylab.org/pjirovecii/) [5]. It was dubbed the ISHAM MLST scheme and consisted of four loci: the mitochondrial large subunit of the rRNA gene region *(mt26S* – but also abbreviated as *mtLSU* in literature), *β-tubulin (β-TUB*), *dihydropteroate synthetase (DHPS),* and internal transcribed region (*ITS1/2*) of the rDNA gene cluster.

Using the ISHAM MLST scheme, genotyping was successfully undertaken to characterize an PCP outbreak at Westmead Hospital's renal transplant ward, successfully identifying the index case, and linking all other cases back to it [5]. There have been no additional PCP outbreaks at

Westmead Hospital since the Phipps findings and recommendations of blanket prophylaxis in renal transplant wards for PCP prevention was issued.

With various MLST schemes for *P. jirovecii* being reported in the literature, the scheme was not being utilized to its full capacity. A reason for this may be due to difficulties identified using the PCR conditions and primer sets described. A fact also seen during in-house testing of positive PCP bronchoalveolar lavage (BAL) samples later in the Medical Mycology Research Group. In addition, the current PCR conditions were complex, with a different primer set and PCR protocol for each individual genetic locus. As such, we aimed to optimize this scheme, by redesigning the PCR protocol for each genetic locus, and define the optimal conditions to amplify the following genetic loci: *β-TUB, DHPS,* ITS1/2, and *mt26S* for *P. jirovecii,* with the aim to generalize them across all four loci.

To show the value of the optimized ISHAM MLST scheme it was then applied to Chilean *Pneumocystis* isolates obtained from HIV positive patients with confirmed PCP infections, investigating their genetic diversity and epidemiological relatedness.

3.2 Methods

3.2.1 *Pneumocystis jirovecii* Samples

P. jirovecii isolates were collected from the Greater Sydney Region, Australia and Santiago, Chile. Two samples from Westmead Hospital, NSW, 80-15-162-4823(C1) and 20-15-162-0618 (C2), were PCP patients, with a high *Pneumocystis* load. These samples were used as positive controls during the optimization process of the existing MLST scheme. In addition, 23 isolated were obtained from PCP positive patients, collected from Chilean hospitals between 2004 and 2014.

Table 3.1. Australian *Pneumocystis jirovecii* samples used as positive controls

Sample ID number	**Date of Isolation**	**Sex and Age**	**Retrieval**	**Type of Respiratory Sample**	**Underlying Conditions**	***P. jirovecii* Presentation**
80-15-162-4823	1/07/17	Male, unknown age	Westmead Hospital, NSW, Australia	BAL	HIV Positive	PCP
20-15-162-0618	17/06/17	Male, unknown age	Westmead Hospital, NSW, Australia	IS	Unknown	PCP

Table 3.2. Chilean *Pneumocystis jirovecii* samples

Sample Number	**Date of Isolation**	**Sex and Age**	**Place of Residence**	**Type of Respiratory Sample**	**Underlying Conditions**	***P. jirovecii*i Presentationn**
515	15/12/04	Male, 34	Santiago, Chile	BAL	Unknown	PCP
1777	17/01/11	Male, 53	Santiago, Chile	BAL	HIV positive	PCP
1794	17/02/11	Male, 23	Santiago, Chile	BAL	HIV positive	PCP
1813	31/03/11	Male, 55	Santiago, Chile	BAL	HIV positive	PCP
1819	6/04/11	Female, 60	Santiago, Chile	BAL	Interstitial Pneumonia	PCP
1833	6/05/11	Male, 30	Santiago, Chile	BAL	HIV positive	PCP
1839	18/05/11	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP
1889	1/08/11	Male, 32	Santiago, Chile	BAL	HIV positive	PCP
1925	5/12/11	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP
2028	3/09/12	Male, 33	Santiago, Chile	BAL	HIV positive	PCP
2061	15/01/13	Male, 45	Santiago, Chile	BAL	HIV positive	PCP
2103	11/12/13	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP
2104	11/12/13	Male, 35	Santiago, Chile	BAL	HIV positive	PCP

2107	30/12/13	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP
2108	7/01/14	Female, unknown age	Santiago, Chile	BAL	HIV positive	PCP
2117	19/02/14	Male, 49	Santiago, Chile	BAL	HIV positive	PCP
2124	19/03/14	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP
2136	16/05/14	Male, 31	Santiago, Chile	BAL	HIV positive	PCP
2156	30/10/14	Female, 40	Santiago, Chile	BAL	HIV positive	PCP
2160	25/11/14	Male, 43	Santiago, Chile	BAL	HIV positive	PCP
2162	7/12/14	Female, 46	Santiago, Chile	BAL	HIV positive	PCP
2164	12/01/15	Male, 54	Santiago, Chile	BAL	HIV positive	PCP
2166	30/01/15	Male, unknown age	Santiago, Chile	BAL	HIV positive	PCP

3.2.2 Extraction and Identification of *Pneumocystis jirovecii*

P. jirovecii was identified using either Gomori Grocott methenamine silver stain or direct immunofluoresce (Meridian Bioscience, Inc.) for all samples [6]. Samples were processed inside a biosafety cabinet using sterile precautions to avoid contamination. They were homogenised with a sterile pipette and a 200 µl aliquot was used for DNA extraction with the QIAamp® DNA Mini kit (Qiagen). *P. jirovecii* was confirmed via PCR of the *mitochondrial large subunit* of the rRNA gene cluster using the primers pAZ102-E and pAZ102-H [7]. Negative controls were included to monitor for cross-contamination during DNA extraction and purification. An internal control using the human *beta-globin* gene was used in each sample to detect false negatives [8]. Each sample was runundituted and as a 1/5 dilution. Amplification products were visualized with ethidium bromide in 2% agarose gels.

3.2.3 The original ISHAM Scheme

The original ISHAM scheme was applied to the Westmead Hospital PCP samples, with three duplicates of each, using the primer conditions (Table 3.3) and PCR protocol (Table 3.4) previously published [5]. Negative controls were included to monitor for cross-contamination during DNA extraction and purification.

Table 3.3. Primers and PCR conditions for the original MLST ISHAM scheme N/A for ITS Round 1 of PRC amplification, as it is the first reaction of the nested PCR, and this product is not run or visualized via gel electrophoresis.

MLST Locus	Nucleotide Sequence of Primers	Product Size (bp)	PCR Conditions
β-TUB	Forward primer: BTubulinF 5´-TCATTAGGTGGTGGAACGGG-3´ Reverse primer: BTubulinR: 5´-ATCACCATATCCTGGATCCG-3´	309	94°C for 3 minutes; followed by 40 cycles: 94°C for 30 seconds, 62°C for 1 minutes, 72°C for 1 minutes; finalised 72°C for 5 minutes
DHPS	Forward primer: DHPS-3 5´-GCGCCTACACATATTATGGCCATTTTAAATC-3´ Reverse primer: DHPS-4 5´-GGAACTTTCAACTTGGCAACCAC-3´	371	95°C for 10 minutes; followed by 10 cycles: 94°C for 15 seconds,72°C for 30 seconds (and decrease 1°C per cycle), 72°C for 15 seconds; then followed by 40 cycles: 92°C for 15 seconds, 62°C for 30 seconds, 72°C for 15 seconds; finalised with 72°C for 5minutes
ITS First round	Forward primer: 1724F2 5´-AGTTGATCAAATTTGGTCATTTAGAG-3´ Reverse primer: ITS2R 5´-CTCGGACGAGGATCCTCGCC-3´	N/A	96°C for 5 minutes; followed by 25 cycles: 94°C for 1 minutes, 60°C for 1 minute, 72°C for 4.5 minutes; finalised with 72°C for 7 minutes
ITS Second round	Forward primer: ITS1F2 5´-CGTAGGTGAACCTGCGGAAGGATC-3´ Reverse primer: ITS2R1 5´-GTTCAGCGGGTGATCCTGCCTG-3´	549	94°C for 5 minutes; followed by 20 cycles: 94°C for 1 minute, 56°C for 1 minutes, 72°C for 4.5 minutes; finalised 72°C for 7 minutes
mt26S	Forward primer: mt26SF 5´-GATGGCTGTTTCCAAGCCCA-3´ Reverse primer: Mt26SR 5´-GTGTACGTTGCAAAGTACTC-3´	307	94°C for 3 minutes; followed by 40 cycles: 94°C for 30 seconds, 52°C for 1 minute, 72°C for 1 minute; finalised 72°C for 5 minutes

Table 3.4. PCR protocols for the optimization of the PCR mix.

Genetic Loci	10 x PCR Buffer	$MgCl_2$	2mM dNTPs	Respective Forward Primer (10ng/µl)	Respective Reverse Primer (10ng/µl)	BioTAQ DNA polymerase (5U/µl)	Sterile water	Genomic DNA (or PCR Product)	
β-TUB	2.5µl	1.5µl	2.5µl	2.5µl	2.5µl	0.5µl	11.375µl	2µl	Each reaction will total to 25 µl
DHPS	2.5µl	1µl	2.5µl	2.5µl	2.5µl	0.5µl	10.5µl	3µl	
ITS1/2 – Round 1	2.5µl	1µl	2.5µl	2µl	2µl	0.5µl	12.5µl	2µl	
ITS1/2 – Round 2	2.5µl	1µl	2.5µl	2µl	2µl	0.5µl	12.5µl	2µl	
mt26S	2.5µl	0.75µl	2.5µl	2.5µl	2.5µl	0.125µl	12.125µl	2µl	

NB: PCR Buffer: Consists of 100 mM Tris-HCL, pH 8.3, 500 mM KCl, 15 mM MgCl2, 0.01 % w/v gelatine. 2 mM dNTPs: Consists of dinucleotide triphosphates, containing 2 mM dATP, dCTP, dGTP, dTTP

3.2.4 PCR Visualization

All PCR amplicons were visualized on 1.5% agarose gel electrophoresis, prepared in 100 ml of TBE buffer (1 M Tris-HCL, 1 M Boric acid, 0.02 M EDTA). 1.5 µl of Ethidium bromide/100 µl agarose solution was added and mixed well. 1 µl of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in water) was added to 5 µl of each PCR sample, and 1 Kb Plus DNA Ladder (Invitrogen) was added for reference. The obtained fragments were separated via electrophoresis at 80 V for 40 minutes, and then visualized under UV illumination on the transilluminator.

3.2.5 DNA Sequencing

Bidirectional sequencing was performed at Macrogen Inc., Seoul, South Korea. The obtained sequences where then assembled and edited using the software package Sequencher ver. 5.4.6 (Gene Codes Corporation). The cleaned-up sequences for each locus were aligned with the program ClustalW [9] part of the software MEGA version 10.1 [10] and compared to reference sequences found on the ISHAM database at http://mlst.mycologylab.org/pjirovecii/. Allele types were named with respect to previously the published nomenclature on the database.

3.2.6 DNA Genotyping Analysis

The visualization of conserved sites in the individual alignments was carried out using BioEdit 7.1.9. The average nucleotide diversity and the number of segregating polymorphic sites were calculated using the software DnaSP ver. 5.10.01 [11]. SNPs were extracted in the statistical environment R [12] using *adegenet* package [13]. Phylogenetic relationships were inferred by the maximum likelihood (ML) method based on the general time reversible (GTR) model with RaxML

(version 7.2.8) using RaxmlGUI 1.1 [14] and the GTRGAMMA option with 1,000 rounds of bootstrap replicates. To test for evidence of recombination based on phylogenetic compatibilities of nearby polymorphic sites along concatenated sequences, the $<\lambda_w$ test [15] implemented in Splitstree 4.10 was used [16]. The test was performed with the default settings of a window size of 100 and k = 2. Single gene ML trees were combined into single file and analysed with Splitstree 4.10 using Consensus Network algorithm (mean distances and thresholds 0.2 and 0.1) [16]. The index of association (IA) is often applied to measure the degree of linkage disequilibrium among different genotypes [17]. Standardized IA and the significance of the null hypothesis of linkage equilibrium (with 10,000 randomizations) were calculated with the program LIAN 3.5 (http://guanine.evolbio.mpg.de/cgi-bin/lian/lian.cgi.pl) using both parametric method and Monte Carlo simulation [18]. The genetic (allelic) diversity was also estimated with LIAN 3.5 software.

3.3 Results

3.3.1 Efficiency of the MLST Scheme

The efficiency of the MLST scheme established by the Australian team at Westmead Hospital was tested, using their original primer designs and PCR protocols, as per Tables 3.3 and 3.4.

The protocols for all four genetic loci in the MSLT scheme were followed exactly as specified with the specific primer sequences originally provided. They were first tested on two different positive PCP samples, 80-15-162-4823 and 20-15-162-0618. None of the four genetic loci, which were targeted in the MLST scheme, yielded positive or acceptable results. The first locus primers used on the *Pneumocystis* samples were the ones for *β-TUB* (Figure 3.1.A). They yielded unsatisfactory results, as they produced thick bright bands, which appeared in all wells, including the negative control, and were well below the expected fragment size. The likely causes of the thickbands are PCR by-products or primer dimers. It also included smeared bands, which just show

nonspecific binding of the primers. Neither are indications of a positive DNA amplification of *P. jirovecii*.

The PCR of the *mt26S* rRNA target locus also yielded unsatisfactory results, producing smeared bands, and thick primer dimer bands at the bottom of all wells (Figure 3.1.B).
The primers for the amplification of *DHPS* gene produced a positive band for one DNA sample of approximately 370 bp in length (Figure 3.1.C). Unfortunately, the other sample had a very long, smeared band instead of a defined band. As such, this result was also unacceptable.

The ITS region is found in multi-copies in the genome, that is why the amplification usually result in high yields. However, to increase sensitivity nested-PCR was adopted to amplify the ITS region. Unfortunately, no positive bands were amplified using this original PCR primers andprotocol. Instead, primer dimers were again noted below each well.

The inadequate and unsuccessful PCR results concluded that the original ISHAM MLST scheme for *P. jirovecii*, along with the suggested primers, and PCR amplification protocol, was not effective in amplifying the four genetic loci from the studied *P. jirovecii* Sydney samples.

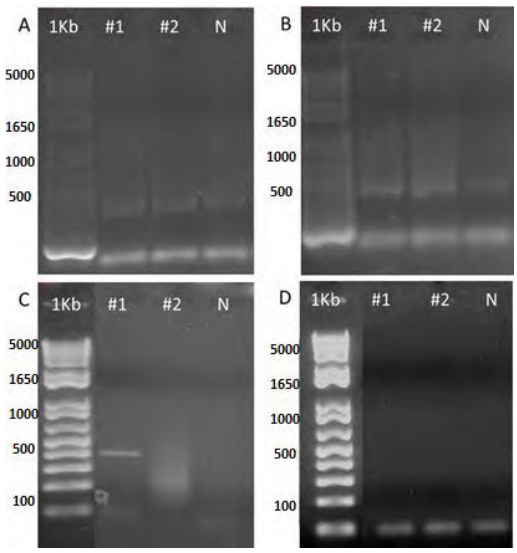


Figure 3.1. Unsuccessful amplification of the four genetic loci of the Australian MLST scheme using the original primers and PCR conditions, PCR results obtained for the two *P. jirovecii* positive Sydney DNA samples, 80-15-162- 4823 (#1) and 20-15-162-0618 (#2). **A)** Amplification of the *β-TUB* locus**. B)** Amplification of the *mt26S* rRNA locus.
C) Amplification of *DHPS* locus. **D)** Amplification of ITS locus. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

3.3.2 Optimization of MLST Primer Sets and PCR Protocol

Due to the unsuccessful PCR amplification using the original ISHAM MLST scheme, PCR protocols were altered to optimize the MLST scheme. This was followed by the design of new PCRprimers to further improve the amplification efficiency.

3.3.2.1 Changes to *β-TUB* PCR Protocol

The primer sets for the *β-TUB* locus remained the same (Table 3.3), but the PCR cycles were increased from 40 to 50 cycles. This produced faint bands, in the expected region of 309 bp (Figure 3.2). Although the bands were present, there were also primer dimer bands below in each well, including the negative control. These results are inconclusive.

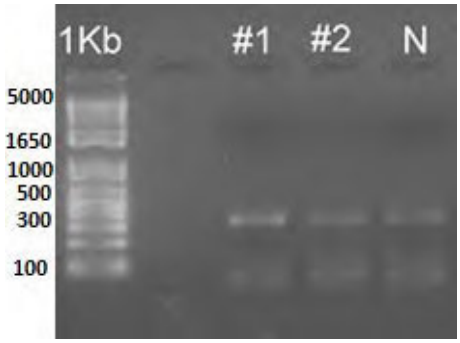


Figure 3.2. Changes in the cycle number of the original PCR protocol yielded unsatisfactory amplification results. PCR Cycles were increased to 50 in an attempt to amplify the *β-TUB* locus from the *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.1.1 *β-TUB* Primer Design

To maximize the amplification of the *β-TUB* locus, one new forward and two new reverse primer was designed.

They are:

PneumoTubForward1: 5´-GCAGGAAATAACTGGGCAAA-3´

PneumoTubReverse1: 5´-CCGAAATGAATGTGATCCCT-3´

PneumoTubReverseR3: 5´-AAACGGCACCATATTAACGG-3´

They were used together in 5 combinations, to be used with the previous primers (Table 3.5). The DNA protocol was changed slightly, as the annealing temperature was lowered to 52°C, while the cycle number were raised to 50.

Table 3.5. *β -TUB* newly designed primer sets.

Combination	Primers	Product size (bp)
***β -TUB* Combination1**	Forward primer: PneumoTubForward1 5´-GCAGGAAATAACTGGGCAAA-3´ Reverse primer: PneumoTubReverse1 5´-CCGAAATGAATGTGATCCCT-3´	656
***β -TUB* Combination2**	Forward primer: BTubulinF 5´TCATTAGGTGGTGGAACGGG-3´ Reverse primer: PneumoTubReverse1 5´-CCGAAATGAATGTGATCCCT-3´	533
***β -TUB* Combination3**	Forward primer: PneumoTubForward1 5´-GCAGGAAATAACTGGGCAAA-3´ Reverse primer: BTubulinR 5´-ATCACCATATCCTGGATCCG-3´	432
***β -TUB* Combination4**	Forward primer: BTubulinF 5´TCATTAGGTGGTGGAACGGG-3´ Reverse primer: PneumoTubReverseR3 5´-AAACGGCACCATATTAACGG-3´	417
***β -TUB* Combination5**	Forward primer: PneumoTubForward1 5´-GCAGGAAATAACTGGGCAAA-3´ Reverse primer: PneumoTubReverseR3 5´-AAACGGCACCATATTAACGG-3´	540

Combination 1 and 4 produced bright, thick bands, at their estimated base pair lengths, without any primer dimers, and no bands in their negative control (Figure 3.3.A and 3.3.D). These were successful amplifications, and either of these two primer sets could confidently be used for the improved MLST scheme. Combination 2 had a smeared band for one sample, and the other sample had a more distinct line, accompanied with a slight smear above it (Figure 3.3.B). Although there were no primer dimers or bands in the negative control, the unspecific binding which caused the

smear deemed the primer combination unsuitable for use. Combination 3 had distinct, bright bands at the correct length, but for sample 20-15-162-0618 there were several non-distinct, faint bands seen above the specific *β-TUB* locus band size (Figure 3.3.C). Combination 5 also had bright, distinct bands, but the fragment size was lower than estimated (Figure 3.3.E). This created a degree of uncertainty in the amplification and primer quality.

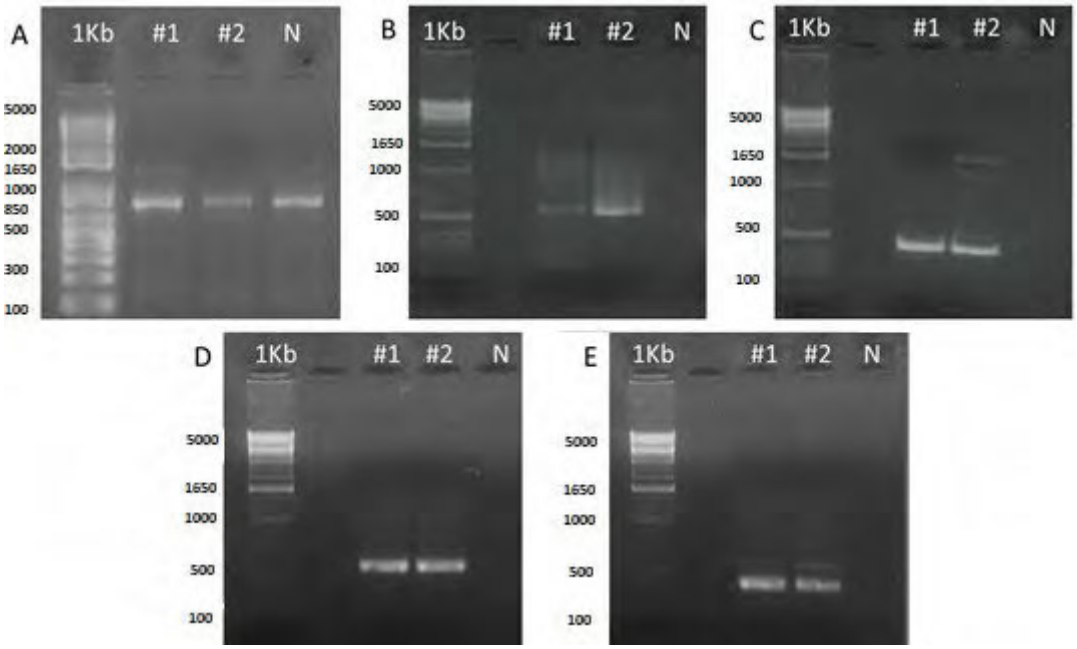


Figure 3.3. Results of the amplification of 5 primer sets for *β-TUB* locus. Primer set combinations from Table 3.5 for *β-TUB* locus were amplified using *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). A) *β-TUB* Combination 1; B) *β-TUB* Combination 2; C) *β-TUB* Combination 3; D) *β-TUB* Combination 4; and E) *β-TUB* Combination 5. The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

3.3.2.2 Changes to *mt26S* PCR Protocol

To improve the performance of the original PCR protocol along with the original *mt26S* primers the cycles of the PCR were increased to 55, from 40. The extended cycles allowed for distinct bands to be created, of the expected fragment length. Although this was a positive result, there were still primer dimers formed, including in the negative control. The PCR mix was then slightly altered, by reducing the amount of reverse and forward primer to 1.5 µl, from 2.5 µl, per 25 µl PCR reaction.

This did not result in the elimination of the primer dimers, and there were still bands in the negative control (Figure 3.4).

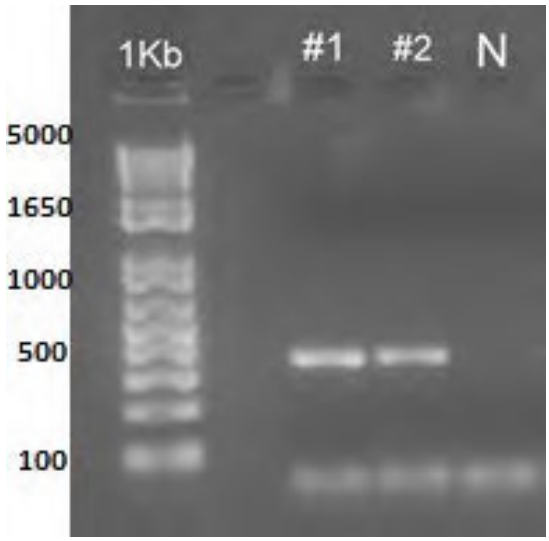


Figure 3.4. Changes to original PCR protocol did not reduce the primer dimers. The PCR cycles were increased to 55, and the PCR mix was adjusted, in attempts to amplify *mt26S* locus on *P. jirovecii* positive Sydney DNA samples, 80- 15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.2.1 *mt26S* Primer Design Round 1

A new primer set was designed to target the *P. jirovecii mt26S* locus. They are:

PneumoLSUForward1: 5´-TCAGGTCGAACTGGTGTACG-3´ and

PneumoLSUReverse2: 5´-TCGGCGAATAGGATTTTCAC-3´

The new primers produced a product size of 601 bp. This primer set was used with the original PCR protocol. Only the cycle number was raised to 55, from 40. There were no distinct bands obtained with this primer set, instead one sample generated a long smear, while the other had just a very faint, non-distinct band, located at a much higher size than expected for the *mt26S* locus (Figure 3.5). The negative control also produced primer dimer bands. As such, this new primer sequence was unacceptable.

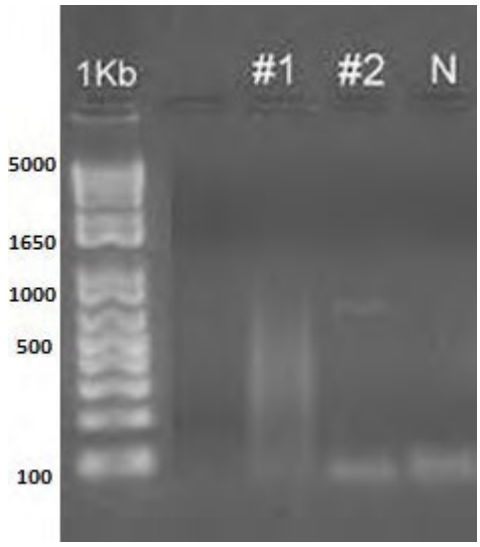


Figure 3.5. The newly designed *mt26S* primers yielded unsuccessful amplification results. The primer set failed to successfully amplify the *mt26S* locus from the *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.2.2 *mt26S* Primer Design Round 2

A new reverse primer was designed, to be used in conjunction with the previous forward primer, PneumoLSUForward1, creating a new primer set. The new primer set was:

PneumoLSUForward1: 5´-TCAGGTCGAACTGGTGTACG-3´ and

PneumoLSUReverse1: 5´-TGTTTCCAAGCCCACTTCTT-3´

This primer set produced a product size of 354bp.

A new PCR amplification protocol was than designed for this primer set: initial denaturation at 94°C for 3 minutes, followed by 50 cycles of: denaturation at 94°C for 30 seconds, annealing at 55°C for 1 minute, and the extension step at 72°C for 1 minute, followed by a final extension stepof 72°C for 5 minutes. PCR amplification under these conditions produced bright, distinct bands forboth DNA samples, and in the correct estimated length of 354 bp (Figure 3.6). There were no other non-distinct bands, primer dimers, or any artifacts seen in the negative band. This positive established that this primer set can confidently be used in the improved MLST scheme.

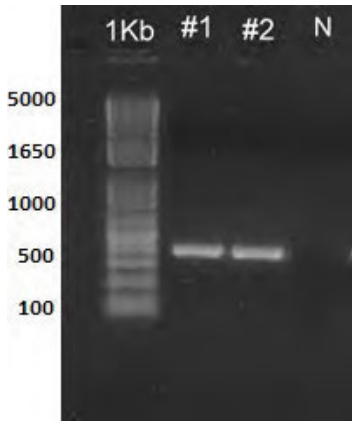


Figure 3.6. Successful *mt26S* amplification using the latest primer design. The primers successfully amplified the *mt26S* locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.3 Changes to *DHPS* PCR Protocol

To improve the amplification efficiency of the *DHPS* locus the PCR amplification protocol was slightly altered. The annealing step during the PCR process, in the second set of cycles was raised to 55°C, but this produced similar results to those seen in Figure 1. There were only non-distinct smeared bands, which were insignificant (Figure 3.7.A).

The PCR protocol was then again changed, and the "10 cycle" step was completely removed. This resulted in non-distinct smears, or faint non-distinct bands, instead of distinct bands (Figure 3.7.B). As a result of those changes this primer set used with either the original protocol or with the altered protocols, was not useful for the MLST scheme.

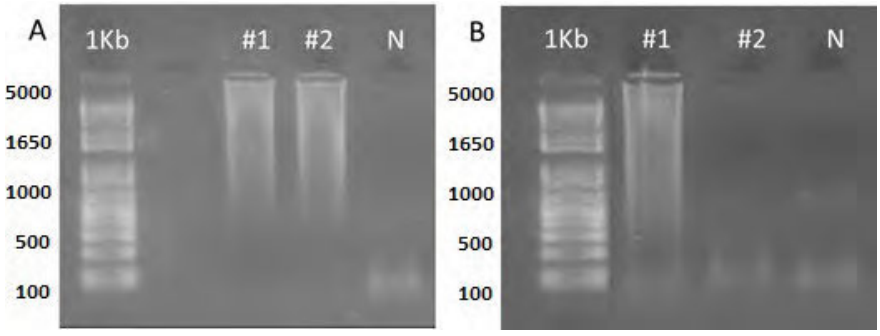


Figure 3.7. Changes to original PCR protocol yielded no *DHPS* amplification products were obtained. **A)** PCR cycle numbers were increased to 55, **B)** The "10 Cycle" step was removed; in attempts to improve the amplification of the *DHPS* locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

3.3.2.3.1 *DHPS* Primer Design Round 1

To further improve the *DHPS* amplification new forward and reverse primers were designed. They are:

PneumoDHPSForward1: 5´-AGCAGTGCCCCAAATCCTAT-3´ and

PneumoDHPSReverse1: 5´-GGCAGTCTACACGGTCTGGT-3´

The new primers produced a product size of 605 bp.

This new primer set was used together with the following new PCR amplification conditions: initial denaturation at 94°C for 3 minutes, followed by 50 cycles of: denaturation at 94°C for 30 seconds, annealing at 60°C for 1 minute, and the extension step at 72°C for 1 minute, followed by a final extension step of 72°C for 5 minutes. The results were similar to those of the previous *DHPS* primers, the bands formed were either very faint and non-distinct or non-distinct smears (Figure 3.8). The new primer set was unsuccessful for the amplification of the *DHPS* locus.

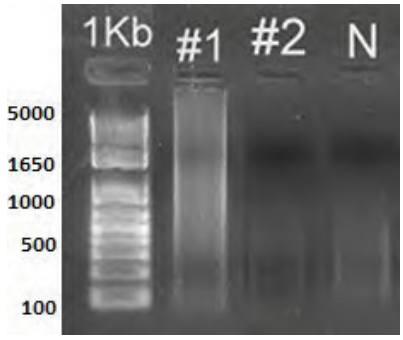


Figure 3.8. The new *DHPS* primer design yielded no amplification results. The newly designed primer set failed to successfully amplify the *DHPS* locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.3.2 *DHPS* Primer Design Round 2

Following the unsuccessful attempts to amplify the *DHPS* locus, another reverse primer wasdesigned, to be used in conjunction with the previously designed forward primer, PneumoDHPSForward1.

The new primer set is:

PneumoDHPSForward1: 5´-AGCAGTGCCCCAAATCCTAT-3´ and

PneumoDHPSReverse2: 5´-GCGCCTACACATATTATGGCCATTTTAAATC-3´

The newly designed primer set produced a product size of 705 bp.

In addition to the new primer the annealing temperature was lowered to 55°C, to better suit the new primers. Bright, distinct bands of the correct fragment size, of 705 bp were obtained. There were no additional non-distinct bands, and the negative control was completely clear (Figure 3.9). This is a successful primer set for the *DHPS* locus that can be used in the improved MLST scheme.

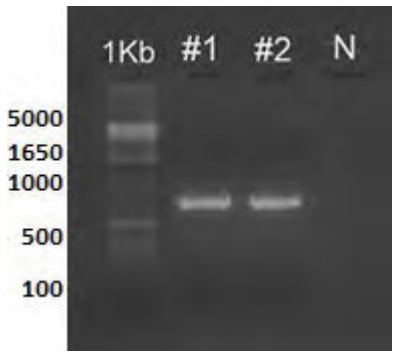


Figure 3.9. Successful *DHPS* amplification using latest primer design. The new primer set successfully amplifies *DHPS* locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE1 Kb Plus DNA Ladder (Invitrogen).

3.3.2.4 Changes to ITS PCR Protocol

To increase the sensitivity of the amplification of the ITS region, a nested-PCR was originally used. As this initial PCR protocol for ITS primers did not yield successful amplifications, the protocol was slightly changed. The cycles for the nested-PCR protocol were raised to 35 for round 1, and 40 for round 2. After these changes, only sample 80-15-162-4823 had a distinct line, but there was no band for sample 20-15-162-0618 (Figure 3.10).

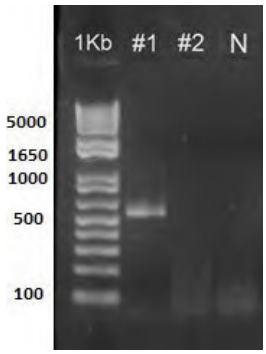


Figure 3.10. Changes to the original PCR protocol yielded no amplification results. The cycle number of the PCR reaction were increased to 35 for Round 1, and 40 for Round 2, in an attempt to amplify the ITS locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen)

3.3.2.4.1 ITS Primer Design Round 1

To standardize the entire MLST scheme, a new primer set, and PCR protocol was developed, to be used in a single step PCR. As such, a new reverse primer was designed, to be used in conjunction with the forward primer from the ITS round 2, ITS1F2 Forward Primer.

The primer set was:

ITS1F2 Forward Primer: 5´-CGTAGGTGAACCTGCGGAAGGATC-3´ and

PneumoITSR: 5´-TCGCCGTTACTAAGGGAATC-3´

This new primer set produced a product size of 608 bp.

The primer set was used under new PCR amplification conditions: initial denaturation at 94°C for 4 minutes, followed by 40 cycles of: denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute, and the extension step at 72°C for 45 seconds, followed by a final extension step of 72°Cfor 7 minutes. The PCR master mix remained the same as those of ITS Round 1. The primer set amplified a distinct band for sample 80-15-162-4823, but there was no distinct band for sample 20- 15-162-0618 (Figure 3.11).

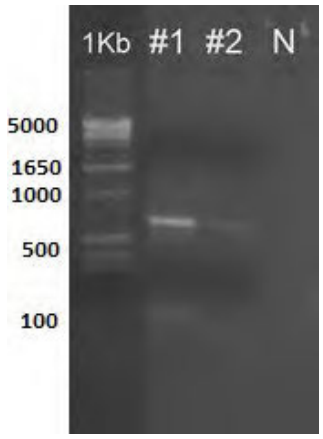


Figure 3.11. The new ITS primer design yielded strong amplification results for sample one but weak results for sample 2. The primer set failed to strongly amplify ITS locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

3.3.2.4.2 ITS Primer Design Round 2

To further improve the PCR efficiency for the ITS amplification a new forward primer wasdesigned to be used with the original reverse primer, PneumoITSR, creating the following primer set:

PneumoITSF: 5´-CCATTGCTGGAAAGTTGATCA-3´ and

PneumoITSR: 5´-TCGCCGTTACTAAGGGAATC-3´

This new primer set produced a product size of 722 bp.

The PCR amplification protocol for this primer set remained the same as used for the amplification products shown in Figure 3.11. Bright, distinct bands were formed using this new primer set, producing the expected 722 bp product size (Figure 3.12). There were no non-distinct bands or smears, and the negative control remained completely clear of any DNA fragments. As such, this new primer set can be confidently used in the MLST scheme for the amplification of the ITS1/2 locus.

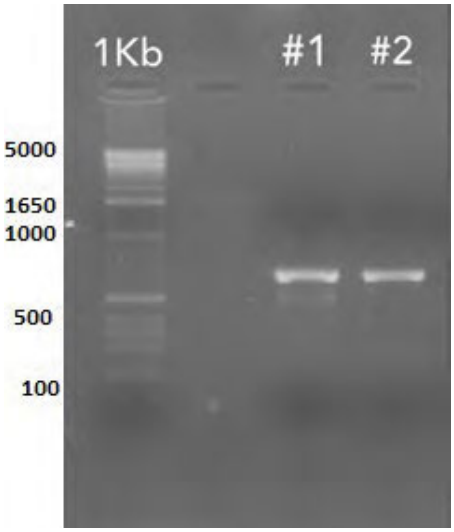


Figure 3.12. Successful ITS amplification using latest primer design. The new primer set successfully amplified the ITS locus from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.3 Final standardization of the PCR Protocol for the improved MLST Scheme

To obtain a more standardized approach for all PCR primers to be used in the new improved MLST scheme (Table 3.9) the master mix and PCR amplification protocol was changed, so they could be applicable to all primers. Judgment of successful amplification was based on which primer set and annealing temperature was able to consistently amplify the DNA to produce clear, thick bands on the gel electrophoresis, and produce reliable DNA sequences. Most successful primer sets withmost successful respective annealing temperature are listed in Table 3.6.

The master mix contained in a total volume of 25 µl per PCR reaction: 2.5 µl 10 x PCR buffer (100 mM Tris-HCL, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$, 0.01 % w/v gelatine), 2 µl 15 mM $MgCl_2$, 2.5 µl 2mM dNTPs (dinucleotide triphosphates, containing 2 mM dATP, dCTP, dGTP, dTTP), 2.5 µl of respective Forward Primer (10 ng/µl), 2.5 µl of respective Reverse Primer (10 ng/µl), 0.3 µl BioTAQ DNA polymerase (5 U/µl), and 2.7 µl of sterile water. Finally, 10 µl of genomic DNA from each clinical fungal sample was added to the PCR reaction tube.

Amplification for all PCR reactions was carried out in the Sensoquest Labcycler thermal cycler using the following protocol: Initial denaturation at 94°C for 3 minutes, followed by 42 cycles of:

denaturation at 94°C for 30 seconds, annealing at 55°C for 45 seconds, and the extension step at 72°C for 45 seconds, followed by a final extension step of 72°C for 7 minutes.

The new master mix and the optimized PCR conditions were applied to both Sydney DNA samples for all four genetic loci of the optimized MLST scheme. All primer sets amplified DNA fragments with their correct size, producing bright, distinct bands (Figure 3.13). There were no non-distinct fragments, and the negative controls remained clear.

Table 3.6. Finalized Primer Sets for the optimized MLST Scheme

MLST Locus	**Primers**	**Product Size (bp)**	**Annealing temperature**
β-TUB	Forward primer: BTubulinF 5´-TCATTAGGTGGTGGAACGGG-3´ Reverse primer: PneumoTubReverseR3 5´- AAACGGCACCATATTAACGG -3´	417	55°C
DHPS	Forward primer: PneumoDHPSF1 5´- AGCAGTGCCCCAAATCCTAT -3´ Reverse primer: PneumoDHPSR2 5´-GCGCCTACACATATTATGGCCATTTTAAATC -3´	705	55°C
mt26S	Forward primer: PneumoLSUForward1 5´- TCAGGTCGAACTGGTGTACG -3´ Reverse primer: PneumoLSUReverse1 5´- TGTTTCCAAGCCCACTTCTT -3´	354	55°C
ITS	Forward primer: PneumoITSF5´- CCATTGCTGGAAAGTTGATCA -3´ Reverse primer: PneumoITSR 5´- TCGCCGTTACTAAGGGAATC -3´	722	55°C

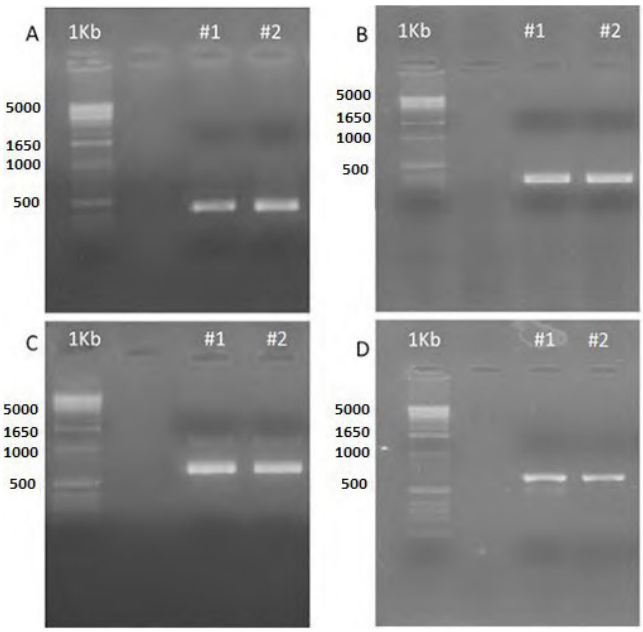


Figure 3.13. Standardized PCR protocol produces successful amplified all four MLST loci. The new primer sets, and optimized PCR protocols successfully amplify all four loci from the two *P. jirovecii* positive Sydney DNA samples, 80-15-162-4823 (#1) and 20-15-162-0618 (#2). **A)** Amplification of the *β-TUB* locus**. B)** Amplification of the *mt26S* rRNA locus. **C)** Amplification of *DHPS* locus. **D)** Amplification of ITS locus. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen).

3.3.4 Proof of concept Application of the optimized MLST scheme to Chilean clinical samples

The improved and optimized MLST scheme was applied to 23 Chilean PCP samples from HIV positive patients. All 23 samples were amplified for all four genetic loci of the optimized MLST scheme. Bright, distinct bands were visualized for all samples, and all negative controls remained clear (Figure 3.14 – 3.17). This confirmed that the loci were correctly amplified. Once sequenced, the samples were able to be analyzed for any epidemiological relatedness between them.

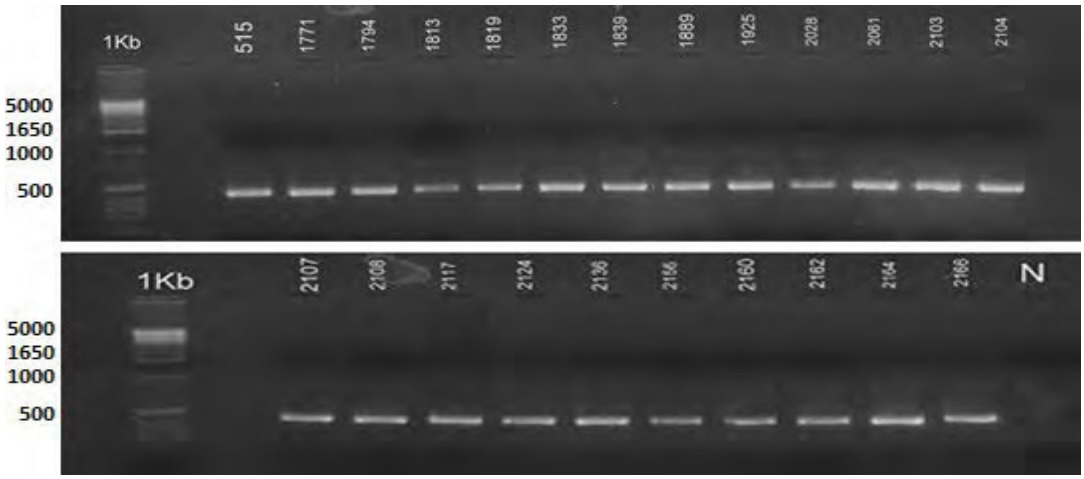


Figure 3.14. The *β-TUB* locus is successfully amplified from all *P. jirovecii* samples from Chile. The new primers and PCR protocol successfully amplified correct size of 417 bp of the *β-TUB* locus from all 23 *P. jirovecii* positive Chilean DNA samples, sample numbers 515 – 2166. The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

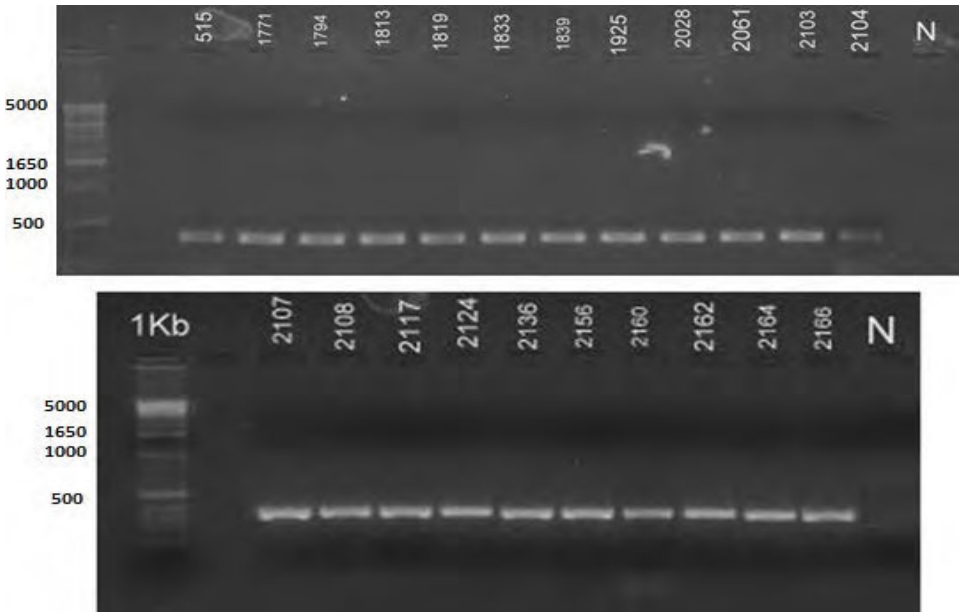


Figure 3.15. The *mt26S* locus is successfully amplified from all *P. jirovecii* samples from Chile. The new primers and PCR protocol successfully amplified the correct size of 354 bp of the *mt26S* locus from all 23 *P. jirovecii* positive Chilean DNA samples, sample numbers 515 – 2166. The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

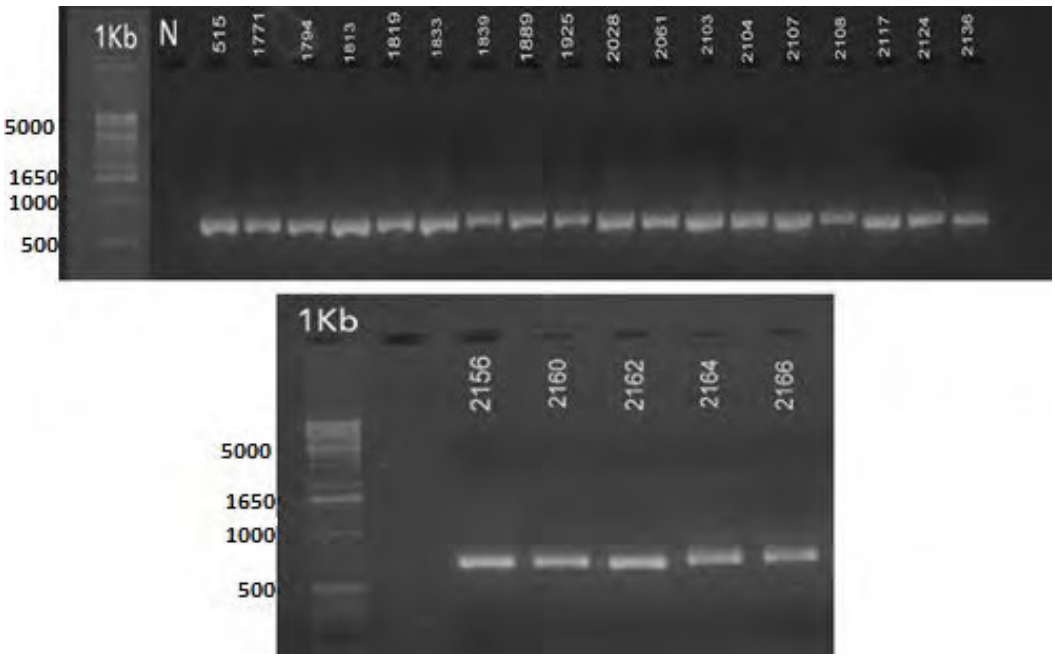


Figure 3.16. The *DHPS* locus is successfully amplified from all *P. jirovecii* samples from Chile. The new primers and PCR protocol successfully amplified the correct size 705 bp of the *DHPS* locus for all 23 *P. jirovecii* positive Chilean DNA samples, sample numbers 515 – 2166. The letter N represents the negative control. Visualized on 1.5% AGE. 1 Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

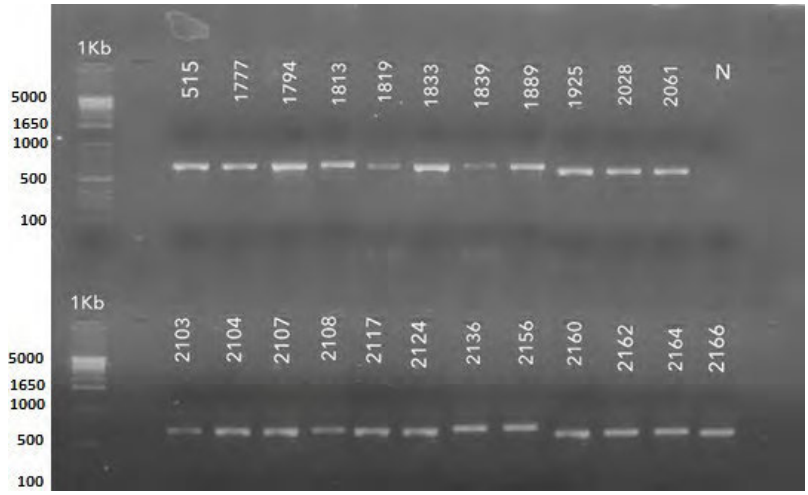


Figure 3.17. The ITS1/2 locus is successfully amplified form all *P. jirovecii* samples from Chile. The new primers and PCR protocol successfully amplified the correct size 722 bp of the ITS locus for all 23 *P. jirovecii* positive Chilean DNA samples, sample numbers 515 – 2166. The letter N represents the negative control. Visualized on 1.5% AGE. 1Kb Plus DNA Ladder (Invitrogen). Image has been modified from original by removing empty wells.

3.3.5 Genetic Analysis of Chilean Isolates

When looking at the analysis of the 23 samples, the ITS locus had 6 allele types identified, the *mt26S* had 4 allele types, the *DHPS* had 3 allele types, and the least diversity was seen in the *β-TUB* locus, with 2 allele types being seen (Table 3.7).

Table 3.7. Allelic diversity and the number of observed alleles in the four MLST loci of the optimized ISHAM MLST scheme.

	Genetic (allelic) diversity	**Genetic (allelic) diversity clone corrected**	**Number of observed alleles**
β -TUB	0.4743	0.4583	2
DHPS	0.3162	0.3417	3
mt26S	0.5929	0.525	4
ITS	0.7273	0.8083	6
Mean (H)	**0.5277 +/- 0.0874**	**0.5333 +/- 0.0992**	

DHPS had one novel allele type (AT3), while *mt26S* had 3 novel allele types (AT4, AT5 and AT6), and ITS had five novel allele types found (AT7, AT8, AT9, AT10, AT11). Dendrograms of the individual loci were constructed, and clusters were evident between all loci. The *DHPS* locus had three district clusters, ITS had two clusters, *mt26S* showed three clusters, and the *β-TUB* had two clear clusters (Figure 3.18).

Further analysis of the DNA sequences was able to determine the genetic diversity between the 23 samples by searching for polymorphic sites (Figure 3.19). The average nucleotide diversity and the number of segregating polymorphic sites were low for all loci. The highest number of polymorphic sites, which was 3, was found in ITS but the nucleotide diversity was the highest in *mt26S*.

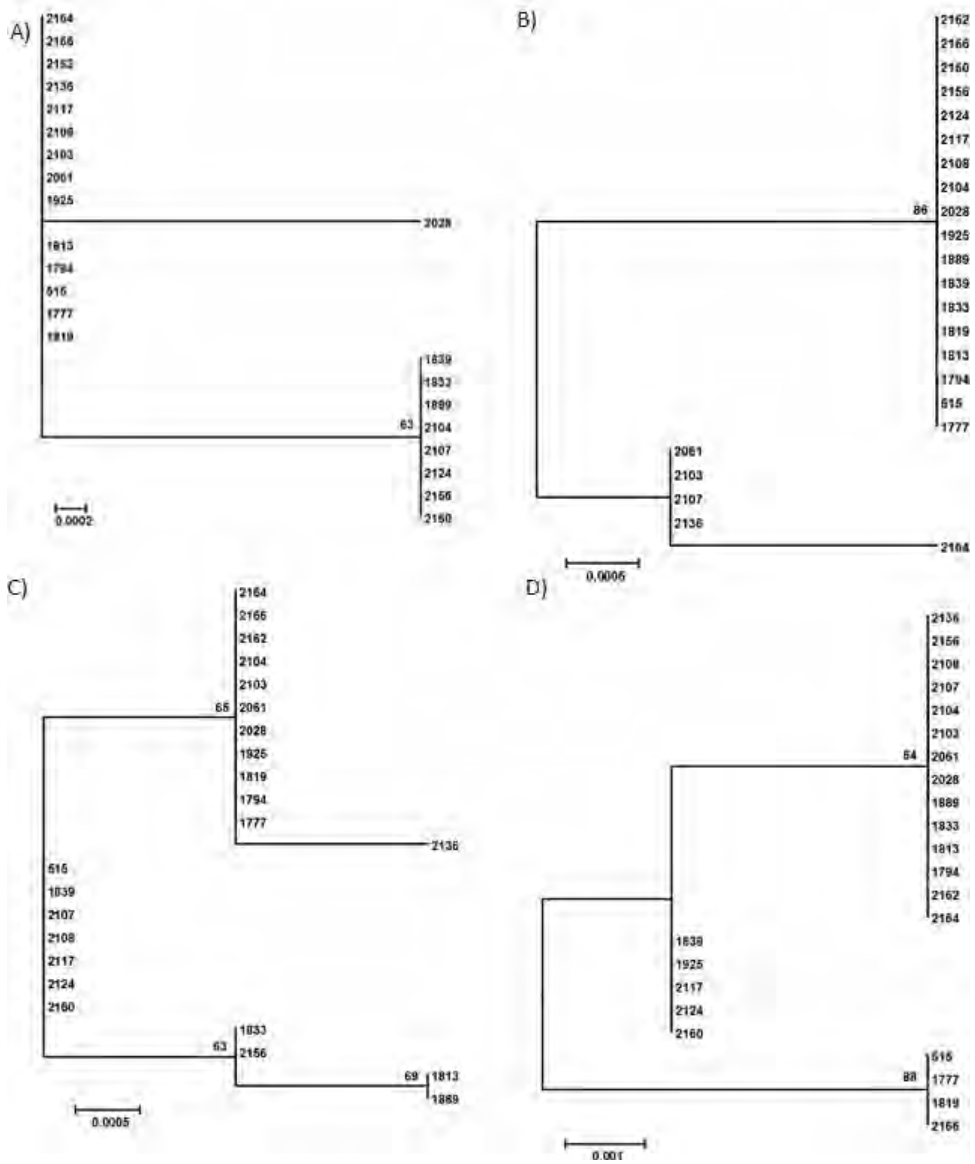


Figure 3.18. Individual allelic phylogenetic trees based of genetic sequences of 23 Chilean *P. jirovecii* strains for loci A) *β-TUB* B) *DHPS* C) ITS1/2, and D) *mt26S.*The numbers given above the branches are frequencies (>60%) with which a given branch appeared in 1,000 bootstrap replications.

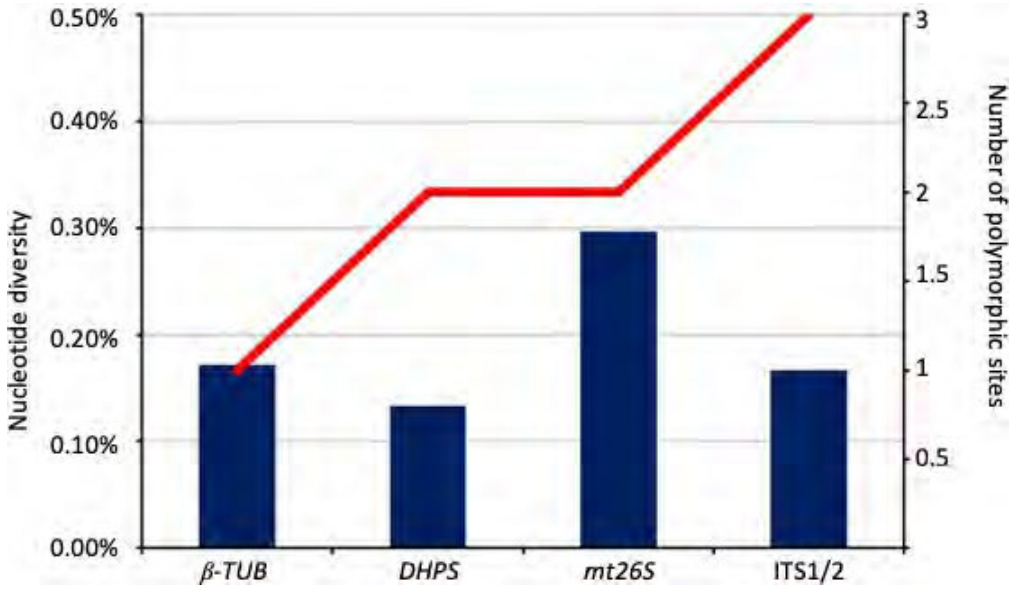


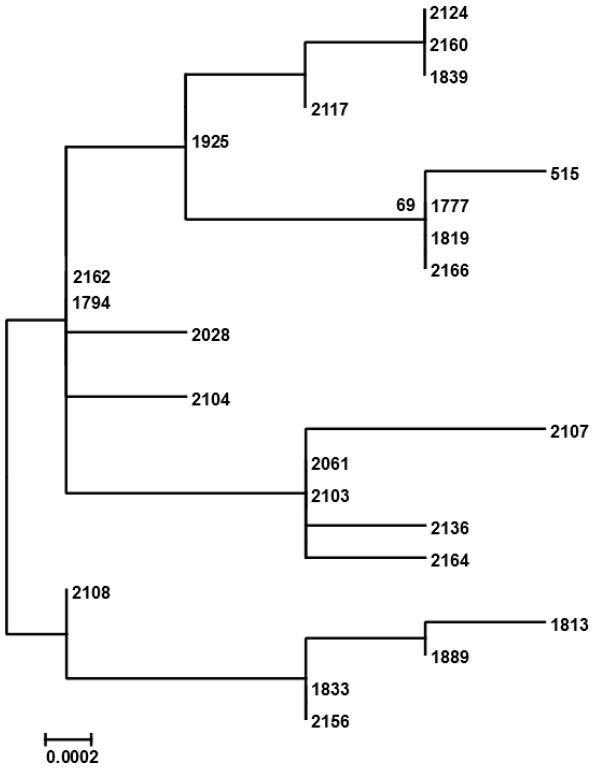
Figure 3.19. Genetic diversity visualized by number of polymorphic sites among 23 Chilean PCP positive DNA samples.

A MLST allele and sequence type scheme table was assembled to determine the variations between the genetic loci analysed. Among the 23 Chilean *P. jirovecii* strains, 16 sequence types were identified (Table 3.8). Strains 1777, 1974 and 2166 have been identified as the same allelic sequence type, number 23. Strains 2103 and 2136 have been identified as the same allelic sequence type, number 20; while strains 1839 and 2124 have been identified as the same allelic sequencetype, number 13. The remaining strains were characterized by unique allelic sequence types. Both a combined phylogenetic tree of all four loci and the consensus network analysis, showed that there were four distinct clustering groups (Figure 3.20).

Table 3.8. The allele type and respective sequence type of 23 Chilean *Pneumocystis jirovecii* isolates, after optimization of the ISHAM MLST scheme.

Isolate number	*-TUB*	*DHPS*	*mt26S*	ITS	Sequence Type
515	1	1	5	7	**26**
1777	1	1	5	8	**23**
1794	1	1	4	8	**21**
1813	1	4	4	8	**28**
1819	1	1	5	8	**23**
1833	2	3	4	9	**14**
1839	2	1	2	8	**13**
1889	2	4	4	9	**17**
1925	1	1	2	8	**22**
2028	1	1	4	9	**19**
2061	1	1	4	11	**24**
2103	1	1	4	10	**20**
2104	2	1	4	9	**15**
2107	2	1	4	10	**16**
2108	1	1	4	7	**25**
2117	1	1	6	8	**27**
2124	2	1	2	8	**13**
2136	1	1	4	10	**20**
2156	2	3	4	9	**14**
2160	2	1	2	8	**13**
2162	1	1	4	8	**21**
2164	1	1	4	2	**18**
2166	1	1	5	8	**23**

A)



B)

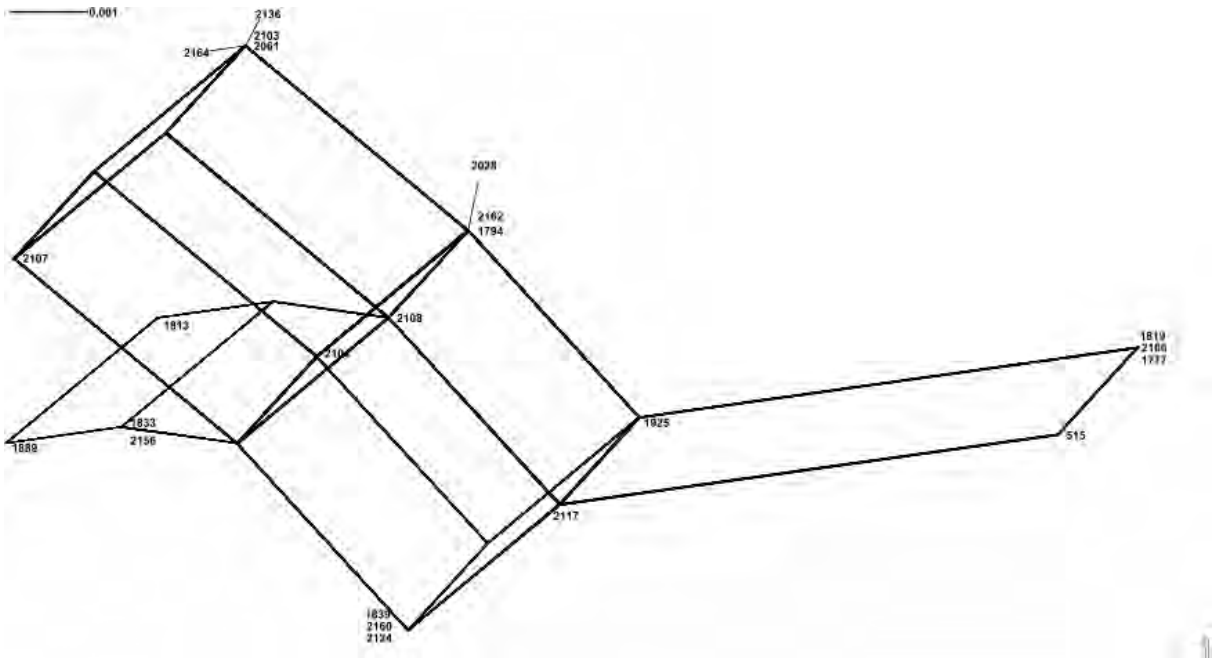


Figure 3.20. A) Phylogenetic tree combining all four genetic MLST loci of ISHAM scheme. The numbers given above the branches are frequencies (>60%) with which a given branch appeared in 1,000 bootstrap replications. **B)** The consensus network analysis obtained from the combined analyses of the single-gene maximum likelihood trees also revealed four clusters as suggested by the ML tree of the combined sequences Be scale indicates mean distance obtained from the analysis of single-gene trees.

3.4 Discussion

The 2011 Phipps *et al.* study demonstrated, genotyping is extraordinarily important for epidemiological investigations of *P. jirovecii* and helping to guide public health approaches towards preventing and handling outbreaks [5]. But, unlike many other fungi, *P. jirovecii* cannot readily be cultured in the laboratory, and therefore the volume of DNA available is always limited, which greatly affects the laboratories' ability to effectively undertake genotyping [19]. As such, an optimalMLST scheme which can successfully amplify and sequence isolates for effective and reliable genotyping is needed, even when faced with minimal DNA volumes.

The original ISHAM MLST scheme targeted the following genetic loci: *β-TUB*, *DHPS*, ITS1/2 and *mt26S*, offering the detection of a high diversity, enabling a reliable distinction of strains. Especially with the inclusion of loci ITS and *mt26S*, which have high discriminatoryindexes [20]. Although the ISHAM MLST scheme had previously worked successfully, there were in-house findings that the scheme was not performing optimally and failing to amplify the loci fromsome of the PCP positive isolates. In addition, the PCR protocol of the scheme was time consumingand complex, which could also cause a low amplification rate. The ISHAM MLST scheme, with its original primers and respective PCR conditions, was retested yielding suboptimal results, even when adding additional PCR cycles, or changing the annealing temperature. This suggested that albeit the ISHAM MLST scheme has a working PCR protocol, it does not always produce reliable results, and when there are lower levels of DNA present in the sample, the amplification is less likely to succeed.

An entire overhaul of the primers and conditions used in the ISHAM MLST scheme was undertaken to improve amplification and DNA sequencing rates when genotyping PCP samples. Several new primer sets were designed for all four loci, which were more specific to target each individual locus, tested in a range of different PRC conditions. The PCR amplification conditions

were significantly changed from the original ones. Additionally, the amplification cycles were also raised, which in some cases was double the number used in the original study. Although this is a considerably high number of cycles for a PCR reaction, it may be a necessity for *Pneumocystis* DNA undergoing PCR amplification, since it is not derived from a pure culture. This study was ableto optimize the PCR protocol and design new primers which improved the PCR amplification and DNA sequencing obtained when genotyping *P. jirovecii* samples (Table 3.6).

Finally, the PCR protocol was standardized across all MLST loci. This generalization of the protocol (including PCR preparation of the master mix) offers an easier and quicker approach to the MLST scheme.

After the MLST scheme was optimized, it was applied to 23 Chilean *Pneumocystis* isolates to test the new optimised ISHAM MLST scheme to detect genetic diversity of the studied samples. The *mt26S* and *DHPS* loci performed with 100% amplification efficiency and produced successful sequences for all studied samples, proving to be highly sensitive and specific for *P. jirovecii*. The two primer sets for the *DHPS* and *mt26S* loci developed in this study should be explored further for their use in the identification of *P. jirovecii* in suspected PCP cases, and possibly also for the use of identifying colonized cases.

Despite multiple attempts of designing optimized PCR primers for the *β-TUB* locus, it still failed the amplification in 3 isolates, and another 6 isolates did not produce adequate DNA sequence quality post amplifications. This can be due to low amplification associated with the samples low genomic DNA concentrations, which is often seen in cases with colonized *P. jirovecii,* or possibly degraded DNA due to aged samples [21]. Additionally, the copy number of a genetic locus in the genome can influences the amplification success [20], which may explain the lower amplification rates obtained for the single copy gene *β-TUB*, in contrast with the multi copy gene *mt26S*, which appears 15 times in the genome [22].

The ITS protocol was changed more significantly, and conventional PCR replaced the older nested PCR method. ITS has been reported multiple times in the literature as having difficulties for amplification [23], and so, not surprisingly performed the worst of the four studied loci. Albeit several studies have reported difficulties with the amplification of ITS, there have also been reports that this can be overcome by the use of nested-PCR [24]. Unfortunately, in this study, nested PCR did not increase the amplification success when compared to the conventional PCR used to amplify the ITS, as the ITS primers for the conventional PCR yielded considerably more satisfactory results.The same 9 isolates which failed sequencing for *β-TUB* also failed sequencing for the ITS, further reinforcing the importance of using high quality DNA and samples which have a high fungal burden to obtain maximum sequencing output. When using samples for genotyping of current or recent outbreaks, this risk can be minimized, especially when obtaining DNA from slightly more invasive sampling methods, such as BAL; or when using samples form biopsies (usually in retrospective genotyping analysis, with the use of autopsied biopsies).

To test the optimized ISHAM MLST scheme for *P. jirovecii* it was applied to 23 Chilean isolates from which all four loci were successfully sequenced. The genetic comparison showed that between the 23 samples there was a low nucleotide diversity and low levels of segregated polymorphic sites in the individual loci. The highest level of polymorphic sites was found in the ITS locus. This was expected, since ITS is derived from the rRNA gene cluster, a region known for its high levels of polymorphisms [25]. ITS currently has also the highest number of unique allele types for *P. jirovecii*, with 120 submitted to GenBank [26]. The nucleotide diversity amongst the 23 isolates was found to be the highest in the *mt26S* loci. This was also an expected outcome, as the *mt26S* locus has previously been widely used in *Pneumocystis* MLST schemes based on its high nucleotide diversity and high discriminatory power, which raised the H-index of MLST schemes, making them more reliable for accurate genotyping [27].

Once MLST allelic profiles were assigned to each *P. jirovecii* sample, and sequence types had been assigned, it was evident that once all loci were combined, there was a high allelic diversity between the samples. Amongst the 23 samples 16 unique sequence types were identified, while only sequence types 13 and 23 had each three samples, which were identical, and sequence types 14 and 20 had only two samples each that were identical. This suggests possible recombination in the population, which was reinforced by a low association index of (IA=0.0128) as supported by the a $<\lambda_w$-test.

Further phylogenetic analysis of the isolates showed four distinct clusters evident in both the dendrogram and the consensus network. The clusters though did not correlate to the years of infection, nor the geographical location. As there were very little patient data available, it is very hard to correlate the sequence types, and to explain the link between some of the samples. This highlights the need for and importance of patient metadata when undertaking genotyping for epidemiological studies. In current outbreaks where patient data and contact tracing is sufficiently obtained, this MLST scheme offers considerate insights into the relationships between the isolates and giving further evidence for inter-human transmissions, without having to work with complex PCR protocols, or unreliable primer sets. Despite the setback of missing patient data, comparing DNA from both past and future infections (where possible) using this MLST scheme, should give a wider snapshot into the Chilean *P. jirovecii* diversity and its evolution over time.

The next step in improving genotyping of *Pneumocystis*, would have been to test the sensitivity of the optimized primers and PCR conditions of the ISHAM MLST scheme on individuals with colonised *P. jirovecii.* Having an effective MLST scheme which can amplify low- level fungal burden would help advance knowledge in the epidemiology of the fungus. This would aid in investigating the role carriers play in nosocomial outbreaks, including patient visitors and hospital staff, potentially impacting hospital prevention and management of PCP in vulnerable

groups and wards. However, due to the low levels of PCP infections currently in the greater Sydney area, it was not possible to test potential carriers and their impact in transmission in this study.

In summary, the considerable optimization of the original ISHAM MLST scheme undertaken herein, developed more specific primers and established a more generalised PCR protocol. This will allow researchers to process samples in a simpler and more time-efficient manner to obtain more reliable sequencing results.

3.5 References

1. Sokulska, M., *et al.*, *Pneumocystis jirovecii—from a commensal to pathogen: clinical and diagnostic review*. Parasitology research, 2015. 114(10): p. 3577-3585.

2. Cano, M.M., *et al.*, *Pneumocystis jiroveci genotypes in the Spanish population.* Clinical infectious diseases, 2004. 39(1): p. 123-128.

3. Monroy-Vaca, E.X., *et al.*, *Genetic diversity of Pneumocystis jirovecii in colonized Cuban infants and toddlers.* Infection, genetics and evolution, 2014. 22: p. 60-66.

4. van Hal, S.J., *et al.*, *Clinical significance and phylogenetic relationship of novel Australian Pneumocystis jirovecii genotypes.* J Clin Microbiol, 2009. 47(6): p. 1818-23.

5. Phipps, L.M., *et al.*, *Nosocomial Pneumocystis jirovecii pneumonia: lessons from a cluster in kidney transplant recipients.* Transplantation, 2011. 92(12): p. 1327-34.

6. Procop, G.W., *et al.*, *Detection of Pneumocystis jiroveci in Respiratory Specimens by Four Staining Methods.* Journal of Clinical Microbiology, 2004. 42(7): p. 3333-3335.

7. Wakefield, A.E., *et al.*, *Amplification of mitochondrial ribosomal RNA sequences from Pneumocystis carinii DNA of rat and human origin.* Molecular and BiochemicalParasitology, 1990. 43(1): p. 69-76.

8. Brancart, F., *et al.*, *Quantitative TaqMan PCR for detection of Pneumocystis jiroveci.* Journal of Microbiological Methods, 2005. 61(3): p. 381-387.

9. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. 22(22): p. 4673-4680.

10. Kumar, S., *et al.*, *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. 35(6): p. 1547-1549.

11. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.* Bioinformatics, 2009. 25(11): p. 1451-1452.

12. Team, R.C., *R: A language and environment for statistical computing*. 2013, Vienna, Austria.

13. Jombart, T. and I. Ahmed, *adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.* Bioinformatics, 2011. 27(21): p. 3070-3071.

14. Silvestro, D. and I. Michalak, *raxmlGUI: a graphical front-end for RAxML.* Organisms Diversity & Evolution, 2012. 12(4): p. 335-337.

15. Bruen, T.C., H. Philippe, and D. Bryant, *A simple and robust statistical test for detecting the presence of recombination.* Genetics, 2006. 172(4): p. 2665-2681.

16. Huson, D.H. and D. Bryant, *Application of phylogenetic networks in evolutionary studies.* Molecular biology and evolution, 2006. 23(2): p. 254-267.

17. Croll, D. and I.R. Sanders, *Recombination in Glomus intraradices, a supposed ancient asexual arbuscular mycorrhizal fungus.* BMC evolutionary biology, 2009. 9(1): p. 13.

18. Haubold, B. and R.R. Hudson, *LIAN 3.0: detecting linkage disequilibrium in multilocus data.* Bioinformatics, 2000. 16(9): p. 847-849.

19. Schildgen, V., *et al.*, *Pneumocystis jirovecii can be productively cultured in differentiated CuFi-8 airway cells.* mBio, 2014. 5(3): p. e01186-14.

20. Maitte, C., *et al.*, *Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* Journal of clinical microbiology, 2013. 51(9): p. 2843-2849.

21. Robin, C., *et al.*, *Molecular demonstration of a Pneumocystis outbreak in stem cell transplant patients: evidence for transmission in the daycare center.* Frontiers inmicrobiology, 2017. 8: p. 700.

22. Valero, C., *et al.*, *Copy Number Variation of Mitochondrial DNA Genes in Pneumocystis jirovecii According to the Fungal Load in BAL Specimens.* 2016. 7.

23. Depypere, M., V. Saegeman, and K. Lagrou, *Typing of Pneumocystis jirovecii by multilocus sequencing: evidence of outbreak?* European Journal of Clinical Microbiology & Infectious Diseases, 2016. 35(6): p. 911-916.

24. Totet, A., *et al.*, *Genotypes at the internal transcribed spacers of the nuclear rRNA operon of Pneumocystis jiroveci in nonimmunosuppressed infants without severe pneumonia.* Journal of clinical microbiology, 2003. 41(3): p. 1173-1180.

25. Lu, J.-J., *et al.*, *Typing of Pneumocystis carinii strains that infect humans based on nucleotide sequence variations of internal transcribed spacers of rRNA genes.* Journal of clinical microbiology, 1994. 32(12): p. 2904-2912.

26. Xue, T., *et al., Genotyping of Pneumocystis jirovecii by Use of a New Simplified Nomenclature System Based on the Internal Transcribed Spacer Regions and 5.8S rRNA Gene of the rRNA Operon.* Journal of Clinical Microbiology, 2019. 57(6).

27. Beard, C.B., *et al., Genetic variation in Pneumocystis carinii isolates from different geographic regions: implications for transmission.* Emerging infectious diseases, 2000.6(3): p. 265.

Chapter 4

Development and Application of the *Pneumocystis jirovecii* ISHAM Consensus MLST Scheme

4.1 Background

Despite the significant improvements of the sequencing success rate achieved by optimizing the original ISHAM MLST scheme in the previous chapter, there were still pitfalls which needed to be addressed. A primary concern which could not be overlooked was, that despite significant optimization and re-design, the scheme was unable to completely genotype all Chilean isolates investigated. Only 23 isolates were included in the final analysis, with 27 isolates missing at least one of the four genetic loci. Most often sequencing failure was observed in the ITS genetic locus, incidentally also the locus within the scheme with the highest discriminatory power, while *DHPS* showed zero discriminatory power, a trend documented within other studies, providing very little phylogenetic insight.

To catapult *P. jirovecii* research and allow for collaborative efforts between research groups and hospitals, a consensus MLST scheme needs to be established and then adopted globally. Unfortunately, due to the unacceptable performance of the ITS and *DHPS* loci, the current ISHAM scheme was not deemed appropriate to be promoted for use as the global consensus scheme until other potential loci having bene tested. To establish a consensus scheme, it was necessary to do a detailed review of the performance of all loci which have been used in the last three decades for *Pneumocystis* genotyping to set up a foundation base for assessing the suitability and effectivity of the loci. Popularity and rate of use within the literature was also an aspect considered in the design, as it was more likely a consensus MLST scheme would be adopted if researchers were familiar with the loci, and already had a substantial degree of established effectivity. Effectivity was measured by the success rate of positive amplification and sufficient high-quality sequencing produced, as well as the discriminatory power produced. A review was undertaken including all *P. jirovecii* MLST genotyping literature, and the established new MLST consensus scheme being then applied to PCP isolates for real-life performance analysis. A consensus MLST scheme consisting of the following

loci *β-TUB, CYB, mt26S* and *SOD,* was put forward alongside the development of a dedicated database. The review was published in the Journal of Fungi in October of 2020 and is included in section 4.2. Once this MLST scheme was established, the scheme was used upon a large set of isolates originating from a single hospital site in Spain to assess the performance of the consensus scheme. The analysis reinforced the schemes high discriminatory power and high success sequencing rate, further solidifying the use of the consensus scheme. This study was submitted and to the journal Medical Mycology and was accepted on the 9th September 2021 and is included in section 4.3. The two publications will boost epidemiological advances in *P. jirovecii* studies by creating a standardized MLST scheme, allowing for the collaboration of all research groups focusing on the fungus. A consensus scheme, designed from the evaluation of all previous schemes put forward, will give future studies the confidence to undertake genotyping investigations without the doubt of which loci to use, and their efficiency. The Spanish cohort study reinforces the capability of the scheme, debuting the efficiency of the scheme for both sequencing and analysis when applied to real-world epidemiological investigations.

4.2 Consensus Multilocus Sequence Typing Scheme for *Pneumocystis jirovecii*

This manuscript was published in the Journal of Fungi in October 2020 (volume 6, issue 4, pages 259-279). L. Pasic is the primary author and contributed to all aspects of this study (Designing and performing the research, analyzing the data, and preparing the manuscript). L. Goterris was involved in the acquisition and analysis of data, and the design of the research. M. Guerrero- Murillo, L. Irinyi and A. Kan were involved in the data analysis. C. Ponce and S. L. Vargas were involved in sample acquisition. M. T. Martin-Gomez and W. Meyer were involved in designing of the research, project administration and funding acquisition. All authors were involved in therevision of the manuscript and support its inclusion in this thesis.

Journal of *Fungi*

MDPI

Review

Consensus Multilocus Sequence Typing Scheme for *Pneumocystis jirovecii*

Lana Pasic [1,2,3,†], Lidia Goterris [4,5,†], Mercedes Guerrero-Murillo [4,5], Laszlo Irinyi [1,2,3], Alex Kan [1,2,3], Carolina A. Ponce [6], Sergio L. Vargas [6], M. Teresa Martin-Gomez [4,5] and Wieland Meyer [1,2,3,7,*]

1 Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Westmead Clinical School, Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney 2006, Australia; lana.pasic@sydney.edu.au (L.P.); laszlo.irinyi@sydney.edu.au (L.I.); alex.kan@sydney.edu.au (A.K.)
2 Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney 2006, Australia
3 Westmead Institute for Medical Research, Westmead 2145, Australia
4 Microbiology Department, Hospital Universitari Vall d'Hebron (VHUH), 08035 Barcelona, Spain; lgoterris@vhebron.net (L.G.); mercedes.guerrero@vhir.org (M.G.-M.); mtmartin@vhebron.net (M.T.M.-G.)
5 Genetics and Microbiology Department, Universitat Autònoma de Barcelona (UAB), 08035 Barcelona, Spain
6 Instituto de Ciencias Biomédicas (ICBM), Facultad de Medicina, Universidad de Chile, Santiago de Chile 8380453, Chile; cponce@med.uchile.cl (C.A.P.); svargas@med.uchile.cl (S.L.V.)
7 Westmead Hospital (Research and Education Network), Westmead 2145, Australia
* Correspondence: wieland.meyer@sydney.edu.au; Tel.: +61-2-86273430
† These authors contributed equally to this work.

Received: 10 September 2020; Accepted: 18 October 2020; Published: 30 October 2020

check for updates

Abstract: *Pneumocystis jirovecii* is an opportunistic human pathogenic fungus causing severe pneumonia mainly in immunocompromised hosts. Multilocus sequence typing (MLST) remains the gold standard for genotyping of this unculturable fungus. However, the lack of a consensus scheme impedes a global comparison, large scale population studies and the development of a global MLST database. To overcome this problem this study compared all genetic regions (19 loci) currently used in 31 different published *Pneumocystis* MLST schemes. The most diverse/commonly used eight loci, *β-TUB*, *CYB*, *DHPS*, ITS1, ITS1/2, *mt26S* and *SOD*, were further assess for their ability to be successfully amplified and sequenced, and for their discriminatory power. The most successful loci were tested to identify genetically related and unrelated cases. A new consensus MLST scheme consisting of four genetically independent loci: *β-TUB*, *CYB*, *mt26S* and *SOD*, is herein proposed for standardised *P. jirovecii* typing, successfully amplifying low and high fungal burden specimens, showing adequate discriminatory power, and correctly identifying suspected related and unrelated isolates. The new consensus MLST scheme, if accepted, will for the first time provide a powerful tool to investigate outbreak settings and undertake global epidemiological studies shedding light on the spread of this important human fungal pathogen.

Keywords: *P. jirovecii*; consensus MLST scheme; *β-TUB*; *CYB*; *mt26S* and *SOD*

1. Introduction

Pneumocystis jirovecii is a major opportunistic pathogen, which can manifest into severe pneumonia, *Pneumocystis* pneumonia (PCP), in immunocompromised patients. PCP can cause interstitial lung disease, along with fever, coughs and dyspnea [1]. The incidence is still relatively high, especially in the developing world, for this underestimated fungus, with reported mortality rates ranging from

10% to 60% [2,3]. Besides causing isolated cases, *P. jirovecii* has been linked to nosocomial outbreaks affecting mainly solid organ transplant recipients with devastating consequences. Besides helping to establish epidemiological links among affected patients, allowing for paths of transmission to be mapped and index cases identified within hospital outbreaks, genotyping is an essential tool to achieve knowledge on more general aspects of the epidemiology of microorganisms.

Multilocus sequence typing (MLST) is currently the preferred standard for genotyping, due to the limited amounts of DNA being available from this unculturable fungus in clinical samples, its reproducibility, inexpensiveness and discriminatory ability [4]. However, unlike many other human pathogenic fungi, *Pneumocystis* has, as yet, no consensus typing scheme, which hinders the ability for global comparison of clinical isolates, since multiple various typing schemes are currently in use [5].

There are 19 coding and non-coding DNA regions which have been explored for *Pneumocystis* genotyping world-wide. The levels of allelic polymorphism fluctuate greatly between the used DNA regions, resulting in varying levels of discriminatory power among 31 schemes currently reported in the literature (see Table 1) [3,5–71]. The lack of standardisation limits the interpretation and comparison of different epidemiological datasets and studies. Molecular typing of *P. jirovecii* is further hampered by the fact that the fungus cannot be cultured in vivo, and hence the volume of DNA available for sequencing analysis is limited [72]. The DNA volume can be further depleted depending on the source and site of extraction, and if the patients are having a mild infection or if they are colonised carriers [73]. Therefore, the ability of a locus to be successfully amplified is equally as important as it is discriminatory power, when deliberating which loci to include in a standardised MLST scheme.

Table 1. Published genetic loci used in *P. jirovecii* genotyping, corresponding multilocus sequence typing (MLST) schemes and obtained allele and sequence types. MLST schemes described are listed chronologically, followed by the respective publications using the specific scheme. The third column indicates the total number of isolates included in the study, followed by the fourth column, which indicates the number of isolates which were able to be successfully sequenced by the study. The fifth column lists the total number of sequence types identified, with the following columns listing the number of allele types found for each genetic locus.

Schemes (Included Loci) and Reference	Country	Total # of Samples	# of Samples Sequenced	# of Sequence Types	Genetic Locus																	
					5.8S	*18S*	*23S*	*26S*	*ITS1*	*ITS2*	*ITS1/2*	*msg*	*mt26S*	*β-TUB*	*TS*	*arom/ EPSP*	*mtSSU*	*DHPS*	*UCS*	*Kex1*	*CYB*	*SOD*	*DHFR*	*TRR1*
Scheme 1 (*5.8S*, *18S*, *26S*, ITS1, ITS2)																						
[6]	USA	15	15	6	1	1		1	2	3	NG												
[7]	GBR	24	24	NG	1	1		1	5	7	NG												
Scheme 2 (ITS1, ITS2, *msg*, *mt26S*)																						
[8]	USA	15	15	NG					NG	NG	NG		NG										
[38]	IND	180	29	NG					NG	NG	NG		NG										
Scheme 3 (*26S*, *β-TUB*, ITS1, *mt26S*)																						
[9]	CHE	11	11	NG				3	3				3	2									
[20]	EUR	212	212	6				6	3				4	3									
[19]	EUR	91	91	28				NG	NG				NG	NG									
[33]	DEU	7	7	2				NG	NG				NG	NG									
[42]	DEU	20	14	NG				2	4				4	1									
[43]	CHE	19	7	1(+)				2	4				3	2									
[50]	DE	18	18	NG				2	3				3	2									
[53]	GBR	670	31	NG				NA	5				4	NA									
[54]	FRA	13	10	3				2	2				3	1									
Scheme 4 (ITS1, ITS2, *mt26S*)																						
[10]	USA	15	15	6					4	5	7		4										
[11]	SWE, FRA	7	7	NG					4	4	NG		3										
[48]	GBR	27	27	NG					NG	NG	2		3										
Scheme 5 (*5.8S*, ITS1, ITS2, *mt26S*, *TS*)																						
[12]	FRA, ITA	20	18	NG	6				3	3	10		4		1								
Scheme 6 (*arom*, ITS1, ITS2, *mt26S*, *mtSSU*)																						
[13]	NLD	6	6	NG					NG	NG	9		3			1	2						
Scheme 7 (ITS1, ITS2)																						
[14]	GLO	207	207	NG					15	14	NG												
[18]	JPN	24	24	NG					11	11	NG												
[31]	ZAF	20	20	NG					11	13	NG												
[49]	SWE	64	64	12					10	12	NG												

Table 1. *Cont.*

Schemes (Included Loci) and Reference	Country	Total # of Samples	# of Samples Sequenced	# of Sequence Types	Genetic Locus																	
					5.8S	*18S*	*23S*	*26S*	*ITS1*	*ITS2*	*ITS1/2*	*msg*	*mt26S*	*β-TUB*	*TS*	*arom/ EPSP*	*mtSSU*	*DHPS*	*UCS*	*Kex1*	*CYB*	*SOD*	*DHFR*	*TRR1*
Scheme 8 (*DHPS*, ITS1, ITS2, *mt26S*)																						
[15]	FRA	14	14	NG					NG	NG	NG		NG					NG						
[39]	PRT, ESP	108	×	NG					12	10	NG		4					4						
[44]	AUS	68	68	NG					8	9	10		2					2						
Scheme 9 (*DHFR*, *DHPS*)																						
[16]	USA	37	37	NG														4				2		
[32]	FRA	33	33	NG														3				2		
[40]	THA	29	18	NG														3				3		
[71]	USA	13	13	NG														2				2		
[70]	COL	98	45	NG														4				2		
Scheme 10 (*DHPS*, *mt26S*)																						
[17]	USA	324	191	14									4					4						
[26]	ESP	255	79	NG									4					4						
[34]	ESP	50	12	NG									4					1						
[35]	USA	442	❋	NG									4					4						
[45]	ESP	60	19	NG									3					1						
[51]	ITA	67	⚑	NG									4					3						
Scheme 11 (*26S*, *β-TUB*, *DHPS*, ITS1, *mt26S*)																						
[21]	USA	22	22	10				2	5				4	2				3						
Scheme 12 (*26S*, *β-TUB*, ITS1, ITS2, *mt26S*)																						
[22]	ITA	25	18	15				4	6	6			4	3										
Scheme 13 (*DHPS*, ITS1, ITS2)																						
[23]	USA	57	37	NG					6	7	NG							3						
[27]	PRT	43	43	NG					15	11	17							2						
[36]	ITA	261	174	NG					NG	NG	9							3						
Scheme 14 (*CYB*, *DHFR*, *DHPS*)																						
[24]	JPN	34	34	NG														4			2		1	
Scheme 15 (*DHPS*, ITS1, ITS2, *mtSSU*)																						
[25]	GBR	2	2	NG					NG	NG	2						1	1						
Scheme 16 (*DHPS*, *mt26S*, *mtSSU*)																						
[28]	GBR, ZWE	51	30	NG									3				4	2						
Scheme 17 (*DHPS*, *mtSSU*, *mt26S*, *SOD*)																						
[29]	GBR	16	16	NG									NG				NG	NG				NG		
[30]	NG	76	76	15									4				4	4				4		

Table 1. *Cont.*

Schemes (Included Loci) and Reference	Country	Total # of Samples	# of Samples Sequenced	# of Sequence Types	Genetic Locus																	
					5.8S	*18S*	*23S*	*26S*	*ITS1*	*ITS2*	*ITS1/2*	*msg*	*mt26S*	*β-TUB*	*TS*	*arom/ EPSP*	*mtSSU*	*DHPS*	*UCS*	*Kex1*	*CYB*	*SOD*	*DHFR*	*TRR1*
Scheme 18 (*DHFR, DHPS*, ITS1, ITS2)																							
[37]	PRT	68	68	NG					NG	NG	19							4					4	
Scheme 19 (*UCS, Kex1*)																							
[39]	PRT	87	35	NG															4	4				
Scheme 20 (*CYB, DHFR, DHPS, mt26S, SOD*)																							
[46]	PRT	102	78	NG									9					6			3	6	6	
Scheme 21 (*β-TUB, CYB, DHFR, DHPS, mt26S, TRR1, TS, SOD*)																								
[47]	PRT	70	▶	48									5	3	⌘			3			7	4	3	⌘
Scheme 22 (*β-TUB, DHPS*, ITS1/2, *mt26S*)																								
[3]	AUS	11	11	2							4		2	2				1						
[55]	AUS	48	48	4							2		2	1				1						
[57]	AUS	7	7	NG							NG		NG	NG				NG						
Scheme 23 (*mt26S, mtSSU*)																								
[52]	FRA, CUB, ESP	75	75	NG									5				3							
Scheme 24 (*26S, β-TUB, CYB, DHFR, DHPS*, ITS1, *mt26S, SOD*)																								
[5]	FRA	23	23	NG				7	9				4	2				1			7	3	3	
Scheme 25 (*26S*, ITS1, ITS2, *mt26S*)																								
[56]	DNK	22	18	3				NG	NG	NG	NG		NG											
Scheme 26 (*CYB*, ITS1, *mt26S, SOD*)																								
[58]	FRA	37	32	NG					NG				NG								NG	NG		
Scheme 27 (*β-TUB, CYB, DHFR, mt26S, SOD*)																								
[59]	BEL	20	20 ˚	NG					NA				4	2							4	2	2	
Scheme 28 (*CYB, mt26S, SOD*)																								
[60]	FRA	24	◎	14									6								5	3		
[61]	FRA	32	32	18									22								14	4		
[62]	FRA	7	7	NG									4								3	2		
[63]	POL	17	◉	8									13								6	2		
[66]	FRA	192	35	17									11								5	2		
[67]	TUR	31	26	6									4								5	3		
[68]	REU, GUF, FRA	47	47	23									5								9	3		
Scheme 29 (*23S, 26S, DHPS*)																								
[64]	BRA	30	30	5			3	2										1						

Table 1. *Cont.*

Schemes (Included Loci) and Reference	Country	Total # of Samples	# of Samples Sequenced	# of Sequence Types	Genetic Locus																	
					5.8S	*18S*	*23S*	*26S*	*ITS1*	*ITS2*	*ITS1/2*	*msg*	*mt26S*	*β-TUB*	*TS*	*arom/ EPSP*	*mtSSU*	*DHPS*	*UCS*	*Kex1*	*CYB*	*SOD*	*DHFR*	*TRR1*
Scheme 30 (*DHFR, DHPS, mt26S*)																								
[65]	IND	37	37	13									3					3					2	
Scheme31 (*CYB, DHPS, mt26S, SOD*)																								
[69]	POL	72	N/A	N/A									3					1			3	2		

Loci in black did not match diversity criteria for further consideration. Loci in blue indicate loci investigated in this study, but not included in the newly proposed MLST scheme. Loci in green indicate loci suggested from this study for the newly proposed consensus MLST scheme. NG = Information not given; NA = No amplification recorded; and 1(+) = Study only listed the sequence types (STs) for test isolates, 5 ST were identified when the controls are included. ✕ = 91 samples amplified for *dihydropteroate synthase* (*DHPS*) and 68 for Internal Transcribed Spacer (ITS); ❋= 100% for *mitochondrial large ribosomal subunit* (*mt26S*) and 53% for *DHPS*; ⚑ = 67 for *mt26S* and 21 for *DHPS*; ▶ = *mt26S* 100%, *cytochrome b oxidase* gene (*CYB*) 61%, *superoxide dismutase* (*SOD*) 74%, *β-tubulin* (*β-TUB*) 49%, *dihydrofolate reductase* gene (*DHFR*) 91%, *DHPS* 96%, *thioredoxin reductase* gene (*Trr1*) and *thymidylate synthase* gene (*TS*) 36%; ⌘ = Null sequence divergence, not included in study further for genotyping; ˚ = ITS no amplification; ◎ = 78% *SOD*, 96.4% *mt26S* and 82.1% *CYB*; ◉ = 17/17 *mt26S* and *CYB*, 5/17 *SOD*. Country codes are according to ISO 3166-1 alpha-3.

The aim of this study was to establish a consensus MLST scheme for *P. jirovecii*, taking into account the previously applied loci, to be used globally for *P. jirovecii* strain typing. Having a global consensus MLST scheme will allow for data exchangeability and comparisons of clinical isolates between laboratories, and the creation of an online world-wide MLST databank for *P. jirovecii* isolates.

2. Investigated Loci and Typing Schemes

To select the most appropriate genetic loci all published loci and respective *P. jirovecii* genotyping schemes have been evaluated. Since 1994 nineteen genetic loci, representing either single or multi-locus genes, have been used in diverse genotyping analyses of *P. jirovecii* (see Table 1). Due to the limited DNA amount extracted from *Pneumocystis* positive clinical samples, the loci were rated based on their previous published amplification and sequencing success rates, as well as the diversity revealed per locus and subsequent ability to discriminate between strains.

2.1. Nuclear rRNA Gene Cluster

Firstly, the multi-copy nuclear rRNA gene cluster was studied. It consists of five components which have been amplified and sequenced previously, including the *18S* rDNA gene, the ITS1 region, the *5.8S* rDNA gene, the ITS2 region and the *26S* rDNA gene [6]. While the rDNA genes are highly conserved, the ITS regions show substantial diversity and as such have been used heavily for identification and genotyping of fungi [12]. The ITS1 and ITS2 regions have demonstrated the highest sequence variation among all loci of the rDNA gene cluster, as evident when sequenced as separate loci or in combination of the two regions (including the *5.8S* gene), using nested-PCR techniques [3]. This has resulted in over 120 unique genotypes for both ITS regions reported and submitted to GenBank [74]. The ITS1 and ITS2 regions were chosen over the other genes in the rDNA gene cluster due to their superior discriminative power.

2.2. Mitochondrial Genes

The mitochondrial large ribosomal subunit (*mt26S*) gene is involved in basic metabolic functions, with 15 copies within the genome [75]. This locus was selected as it has been considered to be a highly informative genetic marker due to its high variability between isolates, as well as being used as the main target world-wide for *P. jirovecii* detection and identification [76]. Another mitochondrial gene also selected was the *cytochrome b* gene, which contains approximately six copies per genome [75]. *Cytochrome b oxidase* gene (*CYB*) has been used widely within MLST genotyping of PCP infections and is increasingly commonly used within European hospitals and laboratories [41]. It has a reported high amplifying and sequencing success rate, although it offers a slightly lower variability than the *mt26S* locus. Although the mitochondrial small subunit (*mtSSU*) rDNA gene has over twice as many copies than *mt26S* gene, it has considerably reported lower variation over the *mt26S* gene, five unique genotypes compared to 25 [75,77]. Additionally, the locus has only appeared in six publications since 1998, when it was first sequenced for use in MLST. For all those reasons the *mtSSU* locus was not further considered for the consensus scheme.

2.3. Nuclear Genes

Finally, three nuclear genes, *β-tubulin* (*β-TUB*), *dihydropteroate synthase* (*DHPS*) and *superoxide dismutase* (*SOD*) were also selected to be included in the study. *β-TUB*, is a single copy gene belonging to the *tubulin* coding gene family, which has been used for *Pneumocystis* identification and genotyping since the 1990's and has been published within MLST schemes over 50 times [78]. Additionally, *β-TUB* has been used as a target locus for PCP diagnosis and is also part of the International Society of Human and Animal Mycology (ISHAM) MLST database (available online http://mlst.mycologylab.org), the only current MLST database specific to *Pneumocystis*, hence warranted further investigation [79]. *SOD* is a single copy gene encoding the production of the enzyme superoxide dismutase, which is commonly used in European studies as an efficient and discriminatory locus for genotyping [29,41].

SOD has a lower variation than *β-TUB*, but several studies have shown that these loci can be used to differentiate between colonised *Pneumocystis* (low burden levels) cases and high burden levels, such as in active PCP cases [80].

The final gene selected to be further investigated was *DHPS*, a highly studied locus due to nonsynonymous, point mutations within codons 55 and 57. These point mutations offer insights to trimethoprim-sulfamethoxazole (TMP-SMZ) resistance, due to the SMZ region of TMP-SMZ [81]. Drug resistances in *P. jirovecii* have been suggested by tracking the variations in *DHPS*, signifying an association between mutations and failure of prophylaxis with sulfa drugs [82]. Due to its high amplification success, along with the ability to give insight to resistance [70], *DHPS* is a key genetic region for many MLST studies. Despite its extensive use, low levels of genetic variation have been reported, with most studies reporting wild-type sequences being detected [83]. *DHPS* is also commonly genotyped together with *DHFR*, the *dihydrofolate reductase* encoding gene, as both enzymes are part of the folic acid pathway [46]. *DHFR* is inhibited by trimethoprim, nevertheless, information regarding the occurrence of mutations in *DHFR* are scarce, and also conflicting [70,82]. Studies have reported varying levels of *DHFR* mutations, from 2% to 60%, with no significant trends within global distributions, limiting its validity to be used for discrimination between samples [16,24,47,84]. Although this locus is still investigated within PCP treatment research, among the literature it shows slightly lower rates of variation than *DHPS*, and so would not add any further information when used alongside *DHPS* within a newly proposed MLST scheme.

The *major surface glycoprotein* (*msg*) gene, upstream conserved sequences (USC) genetic regions of the *kexin-like serine protease* gene (*Kex1*), *thymidylate synthase* gene (*TS*), *thioredoxin reductase* gene (*Trr1*), as well as the *5-enolpyruvylshikimate-3 phosphate synthase* activity (EPSP) region within the *arom* gene were not selected for further analysis within this study. These genetic regions displayed an inability to be adequately sequenced, as seen with the *msg* and *USC* genes [46,85]. The *Trr1*, *TS*, *Kex1* and *arom* gene loci showed a minimal sequence divergence due to being highly conserved housekeeping genes, and, as such, they are not suitable for MLST, as their discriminatory power is too low [13,41].

As a result of this theoretical analysis the following seven genetic loci have been selected for further practical exploration in this study: *β-tubulin* gene, *cytochrome b oxidase* gene, *dihydropteroate synthase* gene, internal transcribed spacer 1 (ITS1), internal transcribed spacer 2 (ITS2), *mitochondrial large ribosomal subunit rRNA* gene (*mt26S*, also known as *LSU-mt26S*) and *superoxide dismutase* gene.

3. Amplification Rate and Variation of Target Loci

Two cohorts of positive *Pneumocystis* DNA samples were independently subjected to amplification of the selected loci. The first cohort contained 44 bronchoalveolar lavage (BAL) samples which were PCP positive and were obtained from patients within Chilean hospitals between 2004 and 2014. The second cohort was composed of 23 oropharyngeal washes (OW) and 63 BAL samples from a single centre in Spain, between 2014 and 2018, with 35 positive PCP cases and 51 colonisations.

P. jirovecii PCP was diagnosed either by Grocott-Gomori's methenamine silver stain or direct immunofluorescence (Meridian Bioscience, Inc.) [86]. Samples were processed inside a biosafety cabinet using sterile precautions to avoid contamination. They were homogenised with a sterile pipette and a 200 µL aliquot was used for DNA extraction with the QIAamp® DNA Blood Mini kit (Qiagen). *P. jirovecii* was confirmed via PCR of the mitochondrial large subunit rRNA using the primers pAZ102-E and pAZ102-H [87]. Negative controls were included to monitor for cross-contamination during DNA extraction and purification. An internal control using the *human β-globin* gene [79] was used in each sample to detect false negatives. Each sample was run undiluted and as a 1/5 dilution.

The samples were subjected to amplification to assess the effectiveness of the loci in a practical setting, within a range of *Pneumocystis* samples. To yield higher success rates the PCR primers and the associated amplification protocols were optimised from previously published conditions. Genes were amplified in volumes of 25 µL per PCR reaction, using 10X buffer (100 mM Tris-HCL, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$ and 0.01% *w/v* gelatine), 50 nM $MgCl_2$, 2 mM dNTPs, 10 ng/µL of each primer,

5 U/μL BIOTAQ DNA (Bioline) polymerase and 10 μL of genomic DNA. The optimised primers and amplification conditions suggested to be used for the new consensus MLST scheme are shown in Table 2. The amplification conditions which have been used to amplify the loci not included in the new consensus MLST scheme are shown in the Supplementary Table S1.

Table 2. Primer information for the loci of the novel MLST scheme.

Locus	Primer Name	Ref.	Nucleotide Sequence	Product Size (Base Pairs)	PCR Conditions
β-TUB	PneumoβTub_F	-	5′-TCATTAGGTGGTGGAACGGG-3′	303	95 °C 3 min; 45 cycles: 94 °C 30 s, 60 °C 45 s, 72 °C 45 s; 72 °C 7 min
	PneumoβTub_R		5′-ATCACCATATCCTGGATCCG-3′		
SOD	MnSODFw	5	5′-GGGTTTAATTAGTCTTTTTAGGCAC-3′	602	
	MnSODRw		5′-CATGTTCCCACGCATCCTAT-3′		
CYB	CytbFw	5	5′-CCCAGAATTCTCGTTTGGTCTATT-3′	579	95 °C 3 min; 45 cycles: 94 °C 30 s, 55 °C 45 s, 72 °C 45 s; 72 °C 7 min
	CytbRw		5′-AAGAGGTCTAAAAGCAGAACCTCAA-3′		
mt26S	PneumoLSU_F	-	5′-TCAGGTCGAACTGGTGTACG-3′	297	
	PneumoLSU_R		5′-TGTTCCAAGCCCACTTCTT-3′		

Bidirectional sequencing was performed at Macrogen Inc., Seoul, South Korea. The obtained sequences where then assembled and edited using the software package Sequencher ver. 5.4.6 (Gene Codes Corporation). The cleaned-up sequences for each locus were aligned with the program CLUSTALW [88] part of the software MEGA version 10.1 [89] and compared to reference sequences listed in Table 2 and Supplementary Table S1. Allele types were named with respect to previously the published nomenclature [5,41].

The obtained amplification and sequencing success rates varied widely, with the nested PCR of the ITS1/2 locus being the lowest, with 2% for the Chilean isolates and only 38% of the Spanish isolates. *DHPS* and *mt26S* loci had the highest, which amplified 83% for the Spanish and 100% for the Chilean isolates and 95% of the Spanish and 100% of the Chilean isolates, respectively. Simple PCR of the ITS1 region with the newly designed primers amplified 47% of the Chilean isolates, although the Spanish were not amplified with this primer. This was then followed by the *β-TUB* locus, which amplified 80% of the Spanish and 78% of the Chilean isolates, the *SOD* locus, which amplified 71% of the Spanish and 91% of the Chilean isolates, and the *CYB* locus, which amplified 94% and 93% of both Spanish and Chilean isolates, respectively. Average amplification rates of the two cohorts are seen in Figure 1. The Fisher exact test statistic value indicated no significant differences between the individual cohort amplification rates for *β-TUB*, *CYB* and *mt26S*, and a significant result at $p < 0.05$ using the Fisher's exact test [90] for the *DHPS*, *SOD* and ITS1/2 loci (for all raw data see Supplementary Table S2).

The genetic loci were then assessed for their ability to discriminate between different strains, as a high variability within individual loci will directly increase the discriminatory value of the consensus scheme. Based on the PCR performance the ITS1/2 and ITS1 loci were not further analysed due to their poor amplification rates. The mitochondrial genes, *mt26S* and *CYB*, were found to have the highest variation, followed by *β-TUB* and *SOD*, as judged from the number of unique allele types obtained. The *DHPS* locus showed the least amount of variation, with the vast majority of alleles corresponding to the wild type and only two variants having been found across the entire collection of samples (Figure 2). A new database of all allele and sequence types has been established at http://mlst.mycologylab.org for the newly proposed consensus *P. jirovecii* MLST scheme.

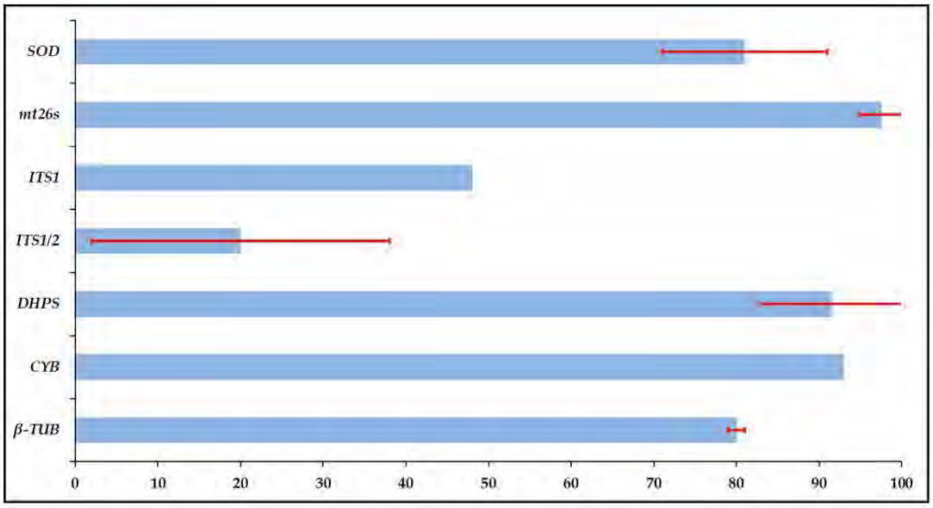


Figure 1. Average amplification and sequencing rates for all targeted genetic loci. Calculations are based on the combined mean amplification and sequencing rates of both cohorts and are expressed as percentages. Error bars indicate the standard deviation.

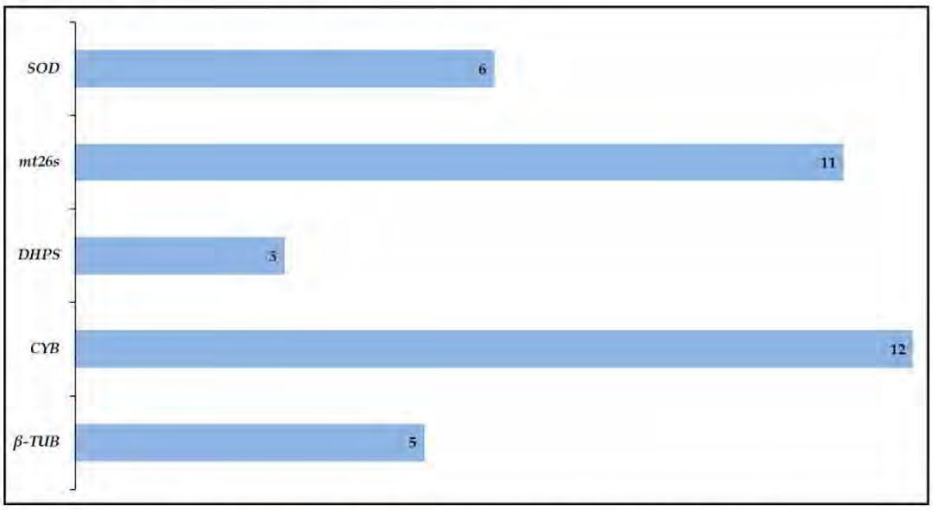


Figure 2. Number of allele types identified amongst all studied isolates.

Based on the above results, showing a superior amplification rate and sequence quality, and high discriminatory power, the following four loci: *β-TUB*, *CYB*, *mt26S* and *SOD* were chosen for inclusion in the proposed new consensus *P. jirovecii* MLST scheme.

4. Case Study: Assessing the Ability to Discriminate between Clinical Isolates

To access the efficiency to identify related and to differentiate between unrelated *P. jirovecii* isolates these four loci were then used to genotype six positive *Pneumocystis* samples representing a potential cluster (two epidemiologically linked isolates) and four independent cases. Allelic profiles were assigned to each sample using the newly developed MLST database available online at

http://mlst.mycologylab.org (Table 3) confirming the suspected to be related isolates and successfully separated all four suspected unrelated isolates (Figure 3). The two suspected related isolates showed identical MLST profiles corresponding both to the ST21. The four suspected unrelated isolates had all unique MLST profiles, ST2, ST7, ST42 and ST44, and where also different from the two related isolates (Table 1 and Figure 3). Cross contamination between samples was ruled out as samples were analysed on different days and results were checked by resequencing a second aliquot.

Table 3. Allele types and sequence types of two related and four unrelated *P. jirovecii* isolates. Colours indicate the different allele types per genetic locus.

Strain Number	Country of Origin	Date of Sample	*β-TUB*	*CYB*	*mt26S*	*SOD*	Sequence Type
HVH21	Spain	Jan 2015	2	1	3	1	ST21
HVH22	Spain	Jan 2015	2	1	3	1	ST21
Case 63	Australia	Dec 2016	1	3	1	4	ST42
Case 71	New Zealand	May 2017	4	2	2	4	ST44
1794	Chile	Feb 2011	2	5	4	1	ST7
2165	Chile	Oct 2014	2	2	4	5	ST2

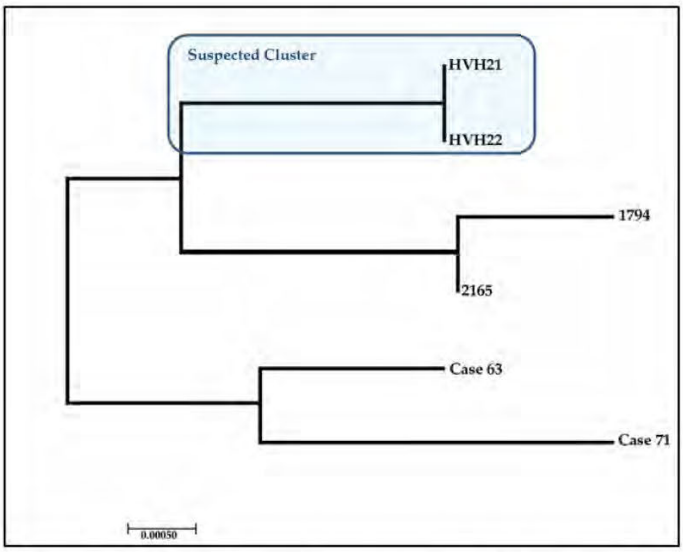


Figure 3. Phylogenetic tree of six *P. jirovecii* isolates (patients from Table 3) used in the case study to show the discriminatory power of the new consensus *P. jirovecii* MLST scheme, obtained by maximum likelihood analysis with the general time reversible (GTR) model with RaxML (version 7.2.8) using RaxmlGUI 1.1 [91], part of the software package MEGA ver. 7.0 [92].

5. Discussion

Genotyping of *P. jirovecii* is vital for the advancement of understanding of the biology, pathogenesis, epidemiology, prophylaxis and treatment regimen of this human pathogen, but more specifically it is vital to help manage, contain and prevent nosocomial clusters. With a rise of nosocomial outbreaks since the early 2000s, hospitals have recorded catastrophic consequences of PCP outbreaks, with up to 83% of reported outbreaks being described within organ transplant wards, as well as in patients with haematological malignancies and connective tissue diseases [93]. Large graft failure and over

50% casualties in wards has been reported from single outbreaks, demonstrating the severity of this underestimated disease [94].

An effective way to investigate epidemiological links is through the creation of transmission maps by combining molecular typing along with studying patient encounters and interactions within the hospital [5]. However, the fact that there are currently 19 genetic regions being used in 31 different typing schemes for *P. jirovecii* limits effective epidemiological studies. The lack of a consensus scheme directly inhibits the ability to compare results, polymorphic strains, and epidemiological data between research centres, and hinders the possibility of establishing global databases and conducting large-scale population studies.

To overcome these limitations, a comprehensive study of all genes and schemes was herein undertaken to assess which loci and which combination of loci would allow for the development of the most practical, efficient and discriminatory scheme. Previous studies looked at the performance of various schemes to suggest a possible consensus scheme, but none have since been formally brought forward as a suggested universal scheme [3]. Maitte et al. 2013 [5], in their review suggested that an eight loci scheme provided the most powerful genotyping results, but this is only possible theoretically due to the limited amount of DNA available from *P. jirovecii* in clinical samples. There are no reliable methods to culture the *P. jirovecii* in vitro, therefore the DNA amount which comes directly from clinical samples is limited [7]. The volume is then further limited depending on the fungal load as well as the source of the specimen [95]. HIV positive patients have high fungal loads, whilst colonised carriers and HIV-negative PCP patients carry lower levels [96]. Bronchoalveolar lavage specimens are the preferred sample, yielding the highest sensitivity due to a greater fungal concentration and also yielding an acceptable negative predictive value [97]. The same studies have shown that the less invasive method, induced sputum (IS), showed comparable levels of fungal burden as BAL, followed by oropharyngeal washes, and then to a lesser extent nasopharyngeal aspirates, and nasal swabs [93,95]. It is therefore vital to have a typing system, consisting of as few genetic loci as possible, but being able to detect and amplify specimens with low levels of fungal burden, as this is often the first limiting step when undergoing *Pneumocystis* genotyping analysis.

On the other side, having too few loci also poses a problem, since lowering the number of loci in a scheme then decreases the discriminatory power and the ability to distinguish between closely related strains. Studies have shown that often schemes with less than three loci do not have enough variation to accurately genotype, as such the Hunter (H)-index has been used in multiple studies to demonstrate the discriminatory power of a scheme [5]. The H-index should not be the only determining factor of an effective scheme as it is highly variable depending on the number of isolates being tested, but it has been a useful tool to help predict the estimated discriminatory power of a scheme [98]. A H-index of 0.95 or higher is considered a suitable cut-off benchmark for MLST schemes, and the review by Maitte et al. 2013 [5] showed that there were no genetic regions which could work individually or paired with another, that would meet this cut-off [19]. As seen in Table 1, there are multiple schemes used in genotyping which are comprised of only two loci, casting a shadow of doubt on their results and generated epidemiological data of these studies, further reinforcing the need for standardisation amongst the *Pneumocystis* scientific community.

As a result, eight genetic regions were explored in this study as potential loci which could be included in a global consensus scheme, by analysing their amplification ability and demonstrating their discriminatory power. The whole ITS region is widely reported to have the most variability within *P. jirovecii* and is highly useful when identifying and genotyping other invasive fungi [74]. Unfortunately, it also has been reported to have a high amplification failure when applied to *P. jirovecii* [59], which was also evident in the current study, as ITS1 alone and ITS1/2 regions showed to have the lowest amplification capability, unable to amplify and successfully sequence more than 50% of the studied isolates for either primer set. Studies have reported that using nested-PCR instead of conventional, single-round PCR can help to overcome the lack of amplification for the ITS region [3,99]. This was not the case in the herein reported study, as the nested-PCR protocol for the ITS1/2 regions performed

considerably worse than single round PCR for the ITS1 region within the Chilean cohort. Due to the lack in amplification success, both in this study and others, no loci from the nuclear rRNA gene cluster were further considered for involvement in a new consensus MLST scheme.

Conversely to the ITS region, *DHPS* amplification was highly successful achieving a global 91.5% amplification success, with minor variations that can be attributed to differences associated to the spectrum of clinical settings covered, ranging from colonisations to infections. *DHPS* is heavily used among *P. jirovecii* genotyping and has been consistently used since the late 1990s in diverse schemes. Despite this, within the reported literature, *DHPS* does not reveal much variation, as such most often wild-types are reported, carrying no informative data to discriminate between strains, showing a H-index of 0 in previous studies [5,100,101]. *DHPS* has demonstrated effectiveness when exploring TMP-SMZ resistance, by showing mutations at the 55 and 57 codons [102]. Population studies using the *DHPS* locus have effectively been able to track additional sulfa pressure in geographical regions with or without widespread use, and as such accurately predict or identify sulfamethoxazole resistance within these regions [103,104]. In this study, only two allele types have been identified, with most of the isolates showing the wildtype, showing regional differences in the distribution of wildtype and allelic variants, subject to further studies. The mutations can be used to give an insight in the widespread use of TMP-SMZ, and for a further comparison between other geographical locations heavily using SMZ but not showing resistance within the community. From a genotyping perspective, the *DHPS* locus, however, does not offer enough variation to be considered as a useful locus within a consensus scheme, and was therefore no longer considered within this study. Despite its drawback to sufficiently discriminate between isolates, researchers should maintain *DHPS* typing for assessing prophylaxis and treatment resistance within populations, and *DHPS* can further be explored for its use as an identifying genetic target in PCP diagnosis due to its successful amplification properties.

Two schemes which use the loci explored in this study and are predominant in the literature are: The official scheme promoted by ISHAM [3], consisting of *β-TUB*, *DHPS*, ITS1 or ITS1/2 and *mt26S*; and a French scheme first proposed by Maitte et al. 2013 [5], consisting of *CYB*, *mt26S* and *SOD*. The ISHAM scheme is highly discriminatory, when applied to both herein studied cohorts it resulted in an H index of >0.98. for both cohorts. The H index was calculated as per Hunder et al. 1988 [98]. Due to different ITS regions explored, ITS1 in the Chilean cohort was able to identify 23 unique sequence types using this scheme, and ITS1/2 in the Spanish cohort attained 14 sequence types. The Maitte et al. 2013 [5] scheme has been reported to have a H-index above 0.95, but this index fell below the threshold when larger sample sizes were considered. H indexes ≤ 0.945 were attained when herein applied to the Chilean and Spanish cohorts, with 33 unique ST detected.

Since the *DHPS* and ITS are not appropriate loci to be used in MLST schemes, the ISHAM scheme could not be promoted for universal application, and neither could the Maitte et al. 2013 [5] scheme, due to inconsistencies with sub-optimal levels of discrimination. *β-TUB*, *CYB*, *mt26S* and *SOD* genetic regions were all individually effective in amplification, but unable to individually discriminate effectively enough, therefore a new consensus MLST scheme comprising of these four loci is herein proposed.

When applied to both cohorts, the new MLST scheme was able to discover a total of 38 unique sequence types, with a combined H-index of 0.975, which is well above the discriminatory cut-off margin for any useful MLST typing scheme.

As this new MLST scheme had not been previously used, the next step was to explore the effectiveness of the scheme to successfully individual *P. jirovecii* isolates. The new MLST scheme was applied to two suspected epidemiologically linked isolates and four isolates for which clinical metadata suggested no relationship, to determine whether it would be discriminatory enough to distinguish the isolates appropriately. The MLST analysis revealed a distinct cluster, showing the genotype ST1, consisting of two patients, named HVH21 and HVH22. The identical sequences of patient HVH21 and HVH22 confirmed the suspected fact of relatedness, which is based on the metadata of the patients, which suggested a possible nosocomial cross-transmission. The cluster involved an HIV positive

colonised patient and a colonised lung transplant recipient, both with a positive BAL, obtained three days apart. Transmission could have taken place at the radiology department, which both had attended on the same day. The other four epidemiologically unlinked isolates were successfully separated from the cluster and revealed each a unique sequence type, ST7, ST13, SR18 and ST21, which was expected as they were obtained from patients from different geographic regions, including Australia, New Zealand and Chile. With the Chilean isolates originating from patients from different Chilean health centres (public and private) three years apart. As such, the newly established consensus MLST scheme has demonstrated its ability to discriminate appropriately between *P. jirovecii* isolates, making it a powerful tool to identify identical strains in settings with many associated cases, such as in an outbreak situation. However, the obtained typing data should always be complemented with a clinical history as good as possible to trace back the origin.

The herein obtained results demonstrated that the new MLST scheme consisting of *β-TUB*, *CYB*, *mt26S* and *SOD* (Table 2) has a much higher amplification rate and an efficient discriminatory power to be applied for genotyping of *P. jirovecii* isolates from clinical samples with high and low fungal burdens, including disease causing and colonising isolates.

Promoting this novel MLST scheme as a global consensus scheme will for the first time standardise MLST for the human pathogen *P. jirovecii* and set up the basis of a substantial improvement in understanding the relationship between clinical *P. jirovecii* isolates. This will allow real-time genotyping of current infected patients and suspected colonised carriers to be now undertaken to improve the understanding of transmissions and the effect colonised carriers have on nosocomial spreads. Further, this will influence public health approaches for preventing nosocomial infections, in especially high-risk patients, such as those recovering from organ transplantation in close approximate in hospital wards.

Since there is no current database, outside of the ISHAM MLST database [3], available online at http://mlst.mycologylab.org/, it is difficult to find all currently published allele types and the lack of standardisation across global centres often cause confusion, even when attempting to compare allele types of a certain locus. The new global database, also placed at http://mlst.mycologylab.org, established herein will improve the nomenclature of allele types and sequence types and makes it easier for researchers and clinician to have one source of information for all genotyping data. The herein standardised MLST scheme will enable the establishment of such a global database, which can be used by all clinical diagnostic and research centres to deposit metadata and sequences, allowing to compare global specimens, something which was not previously possible.

6. Conclusions

In conclusion, this study demonstrates the importance of a consensus MLST scheme for *P. jirovecii* genotyping and the formation of a global database in expanding the understanding of this important human pathogenic fungus. Based on the previous schemes and the evidence in this study, a novel MLST for the genotyping of *P. jirovecii*, consisting of four genetic regions: *β-TUB*, *CYB*, *mt26S* and *SOD* is proposed. This combination of loci maximises the likelihood for amplification and adequate discrimination of isolates over previously used schemes and will aid hospitals in drawing conclusions about interhuman transmission between patients, and hopefully minimise or early detect nosocomial outbreaks.

Supplementary Materials: The following are available online at http://www.mdpi.com/2309-608X/6/4/259/s1, Table S1: Primer information for the initially tested but finally not selected loci. Table S2: Amplification rates of all tested loci for both cohorts.

Author Contributions: Conceptualization, W.M. and M.T.M.-G.; methodology, L.P., L.G., L.I. and C.A.P.; validation, L.P. and L.G.; formal analysis, L.P., L.G., and M.G.-M.; investigation, L.P. and L.G.; resources, W.M. and M.T.M.-G.; data curation, L.P., L.G. and A.K.; writing—original draft preparation, L.P., L.G., M.T.M.-G. and W.M.; writing—review and editing, L.P., L.G., M.G.-M., L.I., A.K., C.A.P., S.L.V., M.T.M.-G. and W.M.; visualization, L.P.; supervision, W.M. and M.T.M.-G.; project administration, W.M.; funding acquisition, W.M., S.L.V. and M.T.M.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a National Health and Medical Research Council of Australia (NH&MRC) grant [no. APP1121936] to W.M., by a National Fund for Science and Technology from Chile (Fondecyt-ANID) grant [no. 1140412] to S.L.V.

Acknowledgments: The authors thank all clinicians and clinical microbiologists who submitted clinical samples to the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meenakshi, K.; Gowtham, R.R.; Usha, K. *Pneumocystis jirovecii* Pneumonia: A Revisit to the Old Malady. *J. Clin. Diagn. Res.* **2019**, *13*, DE01–DE08. [CrossRef]
2. Morris, A.; Lundgren, J.D.; Masur, H.; Walzer, P.D.; Hanson, D.L.; Frederick, T.; Huang, L.; Beard, C.B.; Kaplan, J.E. Current epidemiology of *Pneumocystis* pneumonia. *Emerg. Infect. Dis.* **2004**, *10*, 1713–1720. [CrossRef] [PubMed]
3. Phipps, L.M.; Chen, S.C.-A.; Kable, K.; Halliday, C.L.; Firacative, C.; Meyer, W.; Wong, G.; Nankivell, B.J. Nosocomial *Pneumocystis jirovecii* pneumonia: Lessons from a cluster in kidney transplant recipients. *Transplantation* **2011**, *92*, 1327–1334. [CrossRef] [PubMed]
4. Sokulska, M.; Kicia, M.; Wesołowska, M.; Hendrich, A.B. *Pneumocystis jirovecii*—From a commensal to pathogen: Clinical and diagnostic review. *Parasitol. Res.* **2015**, *114*, 3577–3585. [CrossRef] [PubMed]
5. Maitte, C.; Leterrier, M.; Le Pape, P.; Miegeville, M.; Morio, F. Multilocus sequence typing of *Pneumocystis jirovecii* from clinical samples: How many and which loci should be used? *J. Clin. Microbiol.* **2013**, *51*, 2843–2849. [CrossRef]
6. Lu, J.-J.; Bartlett, M.S.; Shaw, M.M.; Queener, S.F.; Smith, J.W.; Ortiz-Rivera, M.; Leibowitz, M.J.; Lee, C.-H. Typing of *Pneumocystis carinii* strains that infect humans based on nucleotide sequence variations of internal transcribed spacers of rRNA genes. *J. Clin. Microbiol.* **1994**, *32*, 2904–2912. [CrossRef]
7. Tsolaki, A.G.; Miller, R.F.; Underwood, A.P.; Banerji, S.; Wakefield, A.E. Genetic diversity at the internal transcribed spacer regions of the rRNA operon among isolates of *Pneumocystis carinii* from AIDS patients with recurrent pneumonia. *J. Infect. Dis.* **1996**, *174*, 141–156. [CrossRef]
8. Keely, S.P.; Stringer, J.R. Multi-locus genotype switching in *Pneumocystis carinii* sp. f. *hominis*: Evidence for reinfection. *J. Eukaryot. Microbiol.* **1996**, *43*, 50S.
9. Hauser, P.M.; Francioli, P.; Bille, J.; Telenti, A.; Blanc, D.S. Typing of *Pneumocystis carinii* f. sp. *hominis* by single-strand conformation polymorphism of four genomic regions. *J. Clin. Microbiol.* **1997**, *35*, 3086–3091.
10. Keely, S.P.; Stringer, J.R. Sequences of *Pneumocystis carinii* f. sp. *hominis* strains associated with recurrent pneumonia vary at multiple loci. *J. Clin. Microbiol.* **1997**, *35*, 2745–2747.
11. Olsson, M.; Lidman, C.; Latouche, S.; Björkman, A.; Roux, P.; Linder, E.; Wahlgren, M. Identification of *Pneumocystis carinii* f. sp. *hominis* gene sequences in filtered air in hospital environments. *J. Clin. Microbiol.* **1998**, *36*, 1737–1740. [PubMed]
12. Tsolaki, S.; Ortona, E.; Mazars, E.; Margutti, P.; Tamburrini, E.; Siracusano, A.; Guyot, K.; Nigou, M.; Roux, P. Biodiversity of *Pneumocystis carinii hominis*: Typing with different DNA regions. *J. Clin. Microbiol.* **1997**, *35*, 383–387.
13. Tsolaki, A.G.; Beckers, P.; Wakefield, A.E. Pre-AIDS era isolates of *Pneumocystis carinii* f. sp. *hominis*: High genotypic similarity with contemporary isolates. *J. Clin. Microbiol.* **1998**, *36*, 90–93. [PubMed]
14. Lee, C.-H.; Helweg-Larsen, J.; Tang, X.; Jin, S.; Li, B.; Bartlett, M.S.; Lu, J.-J.; Lundgren, B.; Lundgren, J.D.; Olsson, M. Update on *Pneumocystis carinii* f. sp. *hominis* typing based on nucleotide sequence variations in internal transcribed spacer regions of rRNA genes. *J. Clin. Microbiol.* **1998**, *36*, 734–741. [PubMed]
15. Santos, L.D.; Lacube, P.; Latouche, S.; Kac, G.; Mayaud, C.; Marteau, M.; Poirot, J.L.; Maury, E.; Guillot, J.; Roux, P. Contribution of *Dihydropteroate synthase* gene typing for *Pneumocystis carinii* f. sp. *hominis* epidemiology. *J. Eukaryot. Microbiol.* **1999**, *46*, 133S.
16. Ma, L.; Borio, L.; Masur, H.; Kovacs, J.A. *Pneumocystis carinii dihydropteroate synthase* but not *dihydrofolate reductase* gene mutations correlate with prior trimethoprim-sulfamethoxazole or dapsone use. *J. Infect. Dis.* **1999**, *180*, 1969–1978. [CrossRef]

17. Beard, C.B.; Carter, J.L.; Keely, S.P.; Huang, L.; Pieniazek, N.J.; Moura, I.N.; Roberts, J.M.; Hightower, A.W.; Bens, M.S.; Freeman, A.R. Genetic variation in *Pneumocystis carinii* isolates from different geographic regions: Implications for transmission. *Emerg. Infect. Dis.* **2000**, *6*, 265–272. [CrossRef]
18. Hosoya, N.; Takahashi, T.; Wada, M.; Endo, T.; Nakamura, T.; Sakashita, H.; Kimura, K.; Ohnishi, K.; Nakamura, Y.; Mizuochi, T. Genotyping of *Pneumocystis carinii* f. sp. *hominis* isolates in Japan based on nucleotide sequence variations in internal transcribed spacer regions of rRNA genes. *Microbiol. Immunol.* **2000**, *44*, 591–596.
19. Hauser, P.M.; Blanc, D.S.; Sudre, P.; Manoloff, E.S.; Nahimana, A.; Bille, J.; Weber, R.; Francioli, P. Genetic diversity of *Pneumocystis carinii* in HIV-positive and-negative patients as revealed by PCR–SSCP typing. *Aids* **2001**, *15*, 461–466. [CrossRef]
20. Hauser, P.M.; Sudre, P.; Nahimana, A.; Francioli, P.; Group, S. Prophylaxis failure is associated with a specific *Pneumocystis carinii* genotype. *Clin. Infect. Dis.* **2001**, *33*, 1080–1082. [CrossRef]
21. Ma, L.; Kovacs, J.A. Genetic analysis of multiple loci suggests that mutations in the *Pneumocystis carinii* f. sp. *hominis* dihydropteroate synthase gene arose independently in multiple strains. *Antimicrob. Agents Chemother.* **2001**, *45*, 3213–3215. [CrossRef]
22. Volpe, G.; Sbaiz, L.; Avanzini, C.; Caramello, P.; Savoia, D. Genetic diversity of *Pneumocystis carinii* isolated from human immunodeficiency virus-positive patients in Turin, Italy. *J. Clin. Microbiol.* **2001**, *39*, 2995–2998. [CrossRef] [PubMed]
23. Meshnick, S.R.; Hossler, P.A.; Enger, K.S.; Kazanjian, P.; Rest, J.S.; Mindell, D.; Li, B.; Lee, C.; Nimri, L.F.; Carter, J.L. Distribution of *DHPS* mutations among ITS subtypes of *P. carinii* f. sp. *hominis*. *J. Eukaryot. Microbiol.* **2001**. [CrossRef]
24. Takahashi, T.; Kanda, T.; Iwamoto, A. Genetic diversity of drug targets including Dihydropteroate synthase, Dihydrofolate reductase and Cytochrome b, in Pneumocystis carinii f. sp. *hominis* isolates in Japan. *Res. Commun. Mol. Pathol. Pharmacol.* **2002**, *112*, 159–176. [PubMed]
25. Miller, R.F.; Ambrose, H.E.; Novelli, V.; Wakefield, A.E. Probable mother-to-infant transmission of *Pneumocystis carinii* f. sp. *hominis* infection. *J. Clin. Microbiol.* **2002**, *40*, 1555–1557.
26. Montes-Cano, M.A.; De la Horra, C.; Martin-Juan, J.; Varela, J.M.; Torronteras, R.; Respaldiza, N.; Medrano, F.J.; Calderón, E.J. *Pneumocystis jiroveci* genotypes in the Spanish population. *Clin. Infect. Dis.* **2004**, *39*, 123–128. [CrossRef]
27. Matos, O.; Lee, C.-H.; Jin, S.; Li, B.; Costa, M.C.; Gonçalves, L.; Antunes, F. *Pneumocystis jirovecii* in Portuguese immunocompromised patients: Association of specific ITS genotypes with treatment failure, bad clinical outcome and childhood. *Infect. Genet. Evol.* **2003**, *3*, 281–285. [CrossRef]
28. Miller, R.F.; Lindley, A.R.; Ambrose, H.E.; Malin, A.S.; Wakefield, A.E. Genotypes of *Pneumocystis jirovecii* isolates obtained in Harare, Zimbabwe, and London, United Kingdom. *Antimicrob. Agents Chemother.* **2003**, *47*, 3979–3981. [CrossRef] [PubMed]
29. Wakefield, A.E.; Lindley, A.R.; Ambrose, H.E.; Denis, C.-M.; Miller, R.F. Limited asymptomatic carriage of *Pneumocystis jirovecii* in human immunodeficiency virus–infected patients. *J. Infect. Dis.* **2003**, *187*, 901–908. [CrossRef]
30. Miller, R.F.; Lindley, A.R.; Ambrose, H.E.; Aliouat-Denis, C.M.; Wakefield, A.E. Multilocus Genotyping of *Pneumocystis jirovecii* from Adult HIV-Infected Patients with *Pneumocystis* Pneumonia. *J. Eukaryot. Microbiol.* **2003**, *50*, 654–655. [CrossRef] [PubMed]
31. Robberts, F.J.L.; Liebowitz, L.D.; Chalkley, L.J. Genotyping and coalescent phylogenetic analysis of *Pneumocystis jirovecii* from South Africa. *J. Clin. Microbiol.* **2004**, *42*, 1505–1510. [CrossRef] [PubMed]
32. Nahimana, A.; Rabodonirina, M.; Helweg-Larsen, J.; Meneau, I.; Francioli, P.; Bille, J.; Hauser, P.M. Sulfa resistance and *Dihydropteroate synthase* mutants in recurrent *Pneumocystis carinii* pneumonia. *Emerg. Infect. Dis.* **2003**, *9*, 864–867. [CrossRef] [PubMed]
33. Höcker, B.; Wendt, C.; Nahimana, A.; Tönshoff, B.; Hauser, P.M. Molecular evidence of *Pneumocystis* transmission in pediatric transplant unit. *Emerg. Infect. Dis.* **2005**, *11*, 330–332. [CrossRef] [PubMed]
34. Medrano, F.J.; Montes-Cano, M.; Conde, M.; De La Horra, C.; Respaldiza, N.; Gasch, A.; Perez-Lozano, M.J.; Varela, J.M.; Calderon, E.J. *Pneumocystis jirovecii* in general population. *Emerg. Infect. Dis.* **2005**, *11*, 245–250. [CrossRef]

35. Beard, C.B.; Fox, M.R.; Lawrence, G.G.; Guarner, J.; Hanzlick, R.L.; Huang, L.; del Rio, C.; Rimland, D.; Duchin, J.S.; Colley, D.G. Genetic differences in *Pneumocystis* isolates recovered from immunocompetent infants and from adults with AIDS: Epidemiological implications. *J. Infect. Dis.* **2005**, *192*, 1815–1818. [CrossRef]
36. Valerio, A.; Tronconi, E.; Mazza, F.; Cargnel, A.; Fantoni, G.; Atzori, C. *DHPS* -Mutated Isolates of *Pneumocystis jirovecii* from HIV-Infected Individuals: Analysis of Related ITS Genotypes. *J. Eukaryot. Microbiol.* **2006**, *53*, S108–S109. [CrossRef]
37. Costa, M.C.; Esteves, F.; Antunes, F.; Matos, O. Multilocus genotyping of *Pneumocystis jirovecii* in immunocompromised patients: Preliminary results. *J. Eukaryot. Microbiol.* **2006**, *53*, S104–S105. [CrossRef]
38. Gupta, R.; Mirdha, B.R.; Guleria, R.; Mohan, A.; Kabra, S.K.; Kumar, L.; Agarwal, S.K.; Luthra, K. Use of different primer directed sequence amplification by polymerase chain reaction for identification of *Pneumocystis jirovecii* in clinical samples. *Indian J. Chest Dis. Allied Sci.* **2008**, *50*, 321–327.
39. Esteves, F.; Montes-Cano, M.A.; De La Horra, C.; Costa, M.C.; Calderón, E.J.; Antunes, F.; Matos, O. *Pneumocystis jirovecii* multilocus genotyping profiles in patients from Portugal and Spain. *Clin. Microbiol. Infect.* **2008**, *14*, 356–362. [CrossRef]
40. Siripattanapipong, S.; Leelayoova, S.; Mungthin, M.; Worapong, J.; Tan-Ariya, P. Study of *DHPS* and *DHFR* genes of *Pneumocystis jirovecii* in Thai HIV-infected patients. *Med. Mycol.* **2008**, *46*, 389–392. [CrossRef]
41. Esteves, F.; Gaspar, J.; Tavares, A.; Moser, I.; Antunes, F.; Mansinho, K.; Matos, O. Population structure of *Pneumocystis jirovecii* isolated from immunodeficiency virus-positive patients. *Infect. Genet. Evol.* **2010**, *10*, 192–199. [CrossRef] [PubMed]
42. Schmoldt, S.; Schuhegger, R.; Wendler, T.; Huber, I.; Söllner, H.; Hogardt, M.; Arbogast, H.; Heesemann, J.; Bader, L.; Sing, A. Molecular evidence of nosocomial *Pneumocystis jirovecii* transmission among 16 patients after kidney transplantation. *J. Clin. Microbiol.* **2008**, *46*, 966–971. [CrossRef]
43. Gianella, S.; Haeberli, L.; Joos, B.; Ledergerber, B.; Wuethrich, R.-P.; Weber, R.; Kuster, H.; Hauser, P.-M.; Fehr, T.; Mueller, N.J. Molecular evidence of interhuman transmission in an outbreak of *Pneumocystis jirovecii* pneumonia in renal transplant recipients. *Transpl. Infect. Dis.* **2010**, *12*, 1–10. [CrossRef]
44. van Hal, S.J.; Gilgado, F.; Doyle, T.; Barratt, J.; Stark, D.; Meyer, W.; Harkness, J. Clinical significance and phylogenetic relationship of novel Australian *Pneumocystis jirovecii* genotypes. *J. Clin. Microbiol.* **2009**, *47*, 1818–1823. [CrossRef] [PubMed]
45. Montes-Cano, M.A.; Chabe, M.; Fontillon-Alberdi, M.; de la Horra, C.; Respaldiza, N.; Medrano, F.J.; Varela, J.M.; Dei-Cas, E.; Calderon, E.J. Vertical transmission of *Pneumocystis jirovecii* in humans. *Emerg. Infect. Dis.* **2009**, *15*, 125–127. [CrossRef]
46. Esteves, F.; Tavares, A.; Costa, M.C.; Gaspar, J.; Antunes, F.; Matos, O. Genetic characterization of the *UCS* and *Kex1* loci of *Pneumocystis jirovecii*. *Eur. J. Clin. Microbiol. Infect. Dis.* **2009**, *28*, 175–178. [CrossRef] [PubMed]
47. Esteves, F.; Gaspar, J.; Marques, T.; Leite, R.; Antunes, F.; Mansinho, K.; Matos, O. Identification of relevant single-nucleotide polymorphisms in *Pneumocystis jirovecii*: Relationship with clinical data. *Clin. Microbiol. Infect.* **2010**, *16*, 878–884. [CrossRef] [PubMed]
48. Thomas, S.; Vivancos, R.; Corless, C.; Wood, G.; Beeching, N.J.; Beadsworth, M.B. Increasing frequency of *Pneumocystis jirovecii* pneumonia in renal transplant recipients in the United Kingdom: Clonal variability, clusters, and geographic location. *Clin. Infect. Dis.* **2011**, *53*, 307–308. [CrossRef]
49. Beser, J.; Botero-Kleiven, S.; Lebbad, M.; Hagblom, P.; Fernandez, V. A limited number of ITS haplotypes defines the diversity of *Pneumocystis jirovecii* strains in Sweden. *Infect. Genet. Evol.* **2011**, *11*, 948–954. [CrossRef]
50. Pliquett, R.U.; Asbe-Vollkopf, A.; Hauser, P.M.; Presti, L.L.; Hunfeld, K.P.; Berger, A.; Scheuermann, E.H.; Jung, O.; Geiger, H.; Hauser, I.A. A *Pneumocystis jirovecii* pneumonia outbreak in a single kidney-transplant center: Role of cytomegalovirus co-infection. *Eur. J. Clin. Microbiol. Infect. Dis.* **2012**, *31*, 2429–2437. [CrossRef]
51. Dimonte, S.; Berrilli, F.; D'Orazi, C.; D'Alfonso, R.; Placco, F.; Bordi, E.; Perno, C.F.; Di Cave, D. Molecular analysis based on *mtLSU-rRNA* and *DHPS* sequences of *Pneumocystis jirovecii* from immunocompromised and immunocompetent patients in Italy. *Infect. Genet. Evol.* **2013**, *14*, 68–72. [CrossRef] [PubMed]

52. de Armas, Y.; Friaza, V.; Capó, V.; Durand-Joly, I.; Govín, A.; de La Horra, C.; Dei-Cas, E.; Calderón, E.J. Low genetic diversity of *Pneumocystis jirovecii* among Cuban population based on two-locus mitochondrial typing. *Med. Mycol.* **2012**, *50*, 417–420. [CrossRef]
53. Curran, T.; McCaughey, C.; Coyle, P. V *Pneumocystis jirovecii* multilocus genotyping profiles in Northern Ireland. *J. Med. Microbiol.* **2013**, *62*, 1170–1174. [CrossRef]
54. Debourgogne, A.; Favreau, S.; Ladriere, M.; Bourry, S.; Machouart, M. Characteristics of *Pneumocystis* pneumonia in Nancy from January 2007 to April 2011 and focus on an outbreak in nephrology. *J. Mycol. Med.* **2014**, *24*, 19–24. [CrossRef] [PubMed]
55. Nankivell, B.J.; Firacative, C.; Kable, K.; Chen, S.C.-A.; Meyer, W. Molecular epidemiology linking multihospital clusters of opportunistic *Pneumocystis jirovecii* pneumonia. *Clin. Infect. Dis.* **2013**, *57*, 1058–1059. [CrossRef] [PubMed]
56. Rostved, A.A.; Sassi, M.; Kurtzhals, J.A.L.; Sørensen, S.S.; Rasmussen, A.; Ross, C.; Gogineni, E.; Huber, C.; Kutty, G.; Kovacs, J.A. Outbreak of *Pneumocystis* Pneumonia in renal and liver transplant patients caused by genotypically distinct strains of *Pneumocystis jirovecii*. *Transplantation* **2013**, *96*, 834–842. [CrossRef]
57. Chua, K.Y.L.; Halliday, C.L.; Grote, D.; Meyer, W.; Kesson, A.M.; Chen, S.C.A. Colonisation with *Pneumocystis jirovecii* in Australian infants. *Pathol. RCPA* **2015**, *47*, 489–490. [CrossRef] [PubMed]
58. Alanio, A.; Morio, F.; Gits-Muselli, M.; Desnos-Ollivier, M.; Maitte, C.; Bretagne, S. Short tandem repeat genotyping for *P. jirovecii* is more sensitive to mixed genotype than MLST. *J. Mycol. Med.* **2015**, *25*, 220–221. [CrossRef]
59. Depypere, M.; Saegeman, V.; Lagrou, K. Typing of *Pneumocystis jirovecii* by multilocus sequencing: Evidence of outbreak? *Eur. J. Clin. Microbiol. Infect. Dis.* **2016**, *35*, 911–916. [CrossRef]
60. Desoubeaux, G.; Dominique, M.; Morio, F.; Thepault, R.-A.; Franck-Martel, C.; Tellier, A.-C.; Ferrandière, M.; Hennequin, C.; Bernard, L.; Salamé, E. Epidemiological outbreaks of *Pneumocystis jirovecii* pneumonia are not limited to kidney transplant recipients: Genotyping confirms common source of transmission in a liver transplantation unit. *J. Clin. Microbiol.* **2016**, *54*, 1314–1320. [CrossRef]
61. Charpentier, E.; Garnaud, C.; Wintenberger, C.; Bailly, S.; Murat, J.-B.; Rendu, J.; Pavese, P.; Drouet, T.; Augier, C.; Malvezzi, P. Added value of next-generation sequencing for multilocus sequence typing analysis of a *Pneumocystis jirovecii* pneumonia outbreak. *Emerg. Infect. Dis.* **2017**, *23*, 1237–1245. [CrossRef] [PubMed]
62. Vindrios, W.; Argy, N.; Le Gal, S.; Lescure, F.-X.; Massias, L.; Le, M.P.; Wolff, M.; Yazdanpanah, Y.; Nevez, G.; Houze, S. Outbreak of *Pneumocystis jirovecii* infection among heart transplant recipients: Molecular investigation and management of an interhuman transmission. *Clin. Infect. Dis.* **2017**, *65*, 1120–1126. [CrossRef] [PubMed]
63. Sokulska, M.; Kicia, M.; Wesołowska, M.; Piesiak, P.; Kowal, A.; Lobo, M.L.; Kopacz, Ż.; Hendrich, A.B.; Matos, O. Genotyping of *Pneumocystis jirovecii* in colonized patients with various pulmonary diseases. *Med. Mycol.* **2018**, *56*, 809–815. [CrossRef]
64. Ricci, G.; Santos, D.W.; Kovacs, J.A.; Nishikaku, A.S.; de Sandes-Freitas, T.V.; Rodrigues, A.M.; Kutty, G.; Affonso, R.; Silva, H.T.; Medina-Pestana, J.O. Genetic diversity of *Pneumocystis jirovecii* from a cluster of cases of pneumonia in renal transplant patients: Cross-sectional study. *Mycoses* **2018**, *61*, 845–852. [CrossRef]
65. Singh, Y.; Mirdha, B.R.; Guleria, R.; Kabra, S.K.; Mohan, A.; Chaudhry, R.; Kumar, L.; Dwivedi, S.N.; Agarwal, S.K. Genetic polymorphisms associated with treatment failure and mortality in pediatric *Pneumocystosis*. *Sci. Rep.* **2019**, *9*, 1192. [CrossRef] [PubMed]
66. Nevez, G.; Guillaud-Saumur, T.; Cros, P.; Papon, N.; Vallet, S.; Quinio, D.; Minoui-Tran, A.; Pilorgé, L.; de Parscau, L.; Sizun, J. *Pneumocystis* primary infection in infancy: Additional French data and review of the literature. *Med. Mycol.* **2020**, *58*, 163–171. [CrossRef] [PubMed]
67. Sürgeç, E.; Can, H.; Döşkaya, M.; Karakavuk, M.; Atalay Şahar, E.; Değirmenci Döşkaya, A.; Pullukçu, H.; Taşbakan, M.; Sezai Taşbakan, M.; Akyol, D. Genotyping of *Pneumocystis jirovecii* isolates obtained from clinical samples by multilocus sequencing: A molecular epidemiology study conducted in Turkey. *Arch. Microbiol.* **2020**, *202*, 1647–1652. [CrossRef]
68. Le Gal, S.; Toubas, D.; Totet, A.; Dalle, F.; Abou Bacar, A.; Le Meur, Y.; Nevez, G. *Pneumocystis* infection outbreaks in organ transplantation units in France: A Nation-wide survey. *Clin. Infect. Dis.* **2020**, *70*, 2216–2220. [CrossRef]

69. Szydłowicz, M.; Jakuszko, K.; Szymczak, A.; Piesiak, P.; Kowal, A.; Kopacz, Ż.; Wesołowska, M.; Lobo, M.L.; Matos, O.; Hendrich, A.B. Prevalence and genotyping of *Pneumocystis jirovecii* in renal transplant recipients—Preliminary report. *Parasitol. Res.* **2019**, *118*, 181–189. [CrossRef]
70. Muñoz, C.; Zuluaga, A.; Restrepo, A.; Tobón, A.; Cano, L.E.; Gonzalez, A. Molecular diagnosis and detection of *Pneumocystis jirovecii DHPS* and *DHFR* genotypes in respiratory specimens from Colombian patients. *Diagn. Microbiol. Infect. Dis.* **2012**, *72*, 204–213. [CrossRef]
71. Taylor, S.M.; Meshnick, S.R.; Worodria, W.; Andama, A.; Cattamanchi, A.; Davis, J.L.; Yoo, S.D.; Byanyima, P.; Kaswabuli, S.; Goodman, C.D. Low prevalence of *Pneumocystis pneumonia* (PCP) but high prevalence of *Pneumocystis dihydropteroate synthase (dhps)* gene mutations in HIV-infected persons in Uganda. *PLoS ONE* **2012**, *7*, e49991. [CrossRef]
72. Liu, Y.; Fahle, G.A.; Kovacs, J.A. Inability to culture Pneumocystis jirovecii. *MBio* **2018**, *9*, e00939-18. [CrossRef] [PubMed]
73. Roux, P.; Lavrard, I.; Poirot, J.L.; Chouaid, C.; Denis, M.; Olivier, J.L.; Nigou, M.; Miltgen, M. Usefulness of PCR for detection of *Pneumocystis carinii* DNA. *J. Clin. Microbiol.* **1994**, *32*, 2324–2326. [CrossRef] [PubMed]
74. Xue, T.; Ma, Z.; Liu, F.; Du, W.-Q.; He, L.; Ma, L.; An, C.-L. Genotyping of *Pneumocystis jirovecii* by use of a new simplified nomenclature system based on the internal transcribed spacer regions and *5.8S* rRNA gene of the rRNA operon. *J. Clin. Microbiol.* **2019**, *57*, 1–11. [CrossRef]
75. Valero, C.; Buitrago, M.J.; Gits-Muselli, M.; Benazra, M.; Sturny-Leclère, A.; Hamane, S.; Guigue, N.; Bretagne, S.; Alanio, A. Copy number variation of mitochondrial DNA genes in *Pneumocystis jirovecii* according to the fungal load in BAL specimens. *Front. Microbiol.* **2016**, *7*, 1413. [CrossRef] [PubMed]
76. Alanio, A.; Gits-Muselli, M.; Mercier-Delarue, S.; Dromer, F.; Bretagne, S. Diversity of *Pneumocystis jirovecii* during infection revealed by ultra-deep pyrosequencing. *Front. Microbiol.* **2016**, *7*, 733. [CrossRef]
77. Ma, L.; Cissé, O.H.; Kovacs, J.A. A molecular window into the biology and epidemiology of *Pneumocystis* spp. *Clin. Microbiol. Rev.* **2018**, *31*, 1–49. [CrossRef]
78. Nahimana, A.; Francioli, P.; Blanc, D.S.; Bille, J.; Wakefield, A.E.; Hauser, P.M. Determination of the copy number of the nuclear rDNA and *Beta-tubulin* genes of *Pneumocystis carinii* f. sp. *hominis* using PCR multicompetitors. *J. Eukaryot. Microbiol.* **2000**, *47*, 368–372. [CrossRef]
79. Brancart, F.; Rodriguez-Villalobos, H.; Fonteyne, P.-A.; Peres-Bota, D.; Liesnard, C. Quantitative TaqMan PCR for detection of *Pneumocystis jiroveci*. *J. Microbiol. Methods* **2005**, *61*, 381–387. [CrossRef]
80. Morilla, R.; González-Magaña, A.; Friaza, V.; de Armas, Y.; Medrano, F.J.; Calderón, E.; de la Horra, C. Genetic polymorphisms of *Superoxide Dismutase* locus of *Pneumocystis jirovecii* in Spanish population. *Front. Public Health* **2019**, *7*, 292. [CrossRef]
81. Huang, L.; Crothers, K.; Atzori, C.; Benfield, T.; Miller, R.; Rabodonirina, M.; Helweg-Larsen, J. Dihydropteroate synthase gene mutations in *Pneumocystis* and sulfa resistance. *Emerg. Infect. Dis.* **2004**, *10*, 1721–1728. [CrossRef]
82. Nahimana, A.; Rabodonirina, M.; Bille, J.; Francioli, P.; Hauser, P.M. Mutations of *Pneumocystis jirovecii Dihydrofolate reductase* associated with failure of prophylaxis. *Antimicrob. Agents Chemother.* **2004**, *48*, 4301–4305. [CrossRef] [PubMed]
83. Matos, O.; Esteves, F. *Pneumocystis jirovecii* multilocus gene sequencing: Findings and implications. *Future Microbiol.* **2010**, *5*, 1257–1267. [CrossRef]
84. Costa, M.C.; Esteves, F.; Antunes, F.; Matos, O. Genetic characterization of the *Dihydrofolate reductase* gene of *Pneumocystis jirovecii* isolates from Portugal. *J. Antimicrob. Chemother.* **2006**, *58*, 1246–1249. [CrossRef] [PubMed]
85. Sunkin, S.M.; Stringer, J.R. Translocation of *surface antigen* genes to a unique telomeric expression site in *Pneumocystis carinii*. *Mol. Microbiol.* **1996**, *19*, 283–295. [CrossRef]
86. Procop, G.W.; Haddad, S.; Quinn, J.; Wilson, M.L.; Henshaw, N.G.; Reller, L.B.; Artymyshyn, R.L.; Katanik, M.T.; Weinstein, M.P. Detection of *Pneumocystis jiroveci* in respiratory specimens by four staining methods. *J. Clin. Microbiol.* **2004**, *42*, 3333–3335. [CrossRef] [PubMed]
87. Wakefield, A.E.; Pixley, F.J.; Banerji, S.; Sinclair, K.; Miller, R.F.; Moxon, E.R.; Hopkin, J.M. Amplification of mitochondrial ribosomal RNA sequences from *Pneumocystis carinii* DNA of rat and human origin. *Mol. Biochem.* **1990**, *43*, 69–76. [CrossRef]

88. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]
89. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
90. Fisher, M.E.; Scesney, P.E. Visibility of Critical-Exponent Renormalization. *Phys. Rev. A* **1970**, *3*, 825. [CrossRef]
91. Silvestro, D.; Michalak, I. raxmlGUI: A graphical front-end for RAxML. *Org. Divers. Evol.* **2012**, *12*, 335–337. [CrossRef]
92. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol. Biol. Evol.* **2015**, *33*, 1870–1874. [CrossRef]
93. De Boer, M.G.J.; de Fijter, J.W.; Kroon, F.P. Outbreaks and clustering of *Pneumocystis* pneumonia in kidney transplant recipients: A systematic review. *Med. Mycol.* **2011**, *49*, 673–680. [PubMed]
94. Kim, J.E.; Han, A.; Lee, H.; Ha, J.; Kim, Y.S.; Han, S.S. Impact of *Pneumocystis jirovecii* pneumonia on kidney transplant outcome. *BMC Nephrol.* **2019**, *20*, 212. [CrossRef] [PubMed]
95. Alanio, A.; Hauser, P.M.; Lagrou, K.; Melchers, W.J.G.; Helweg-Larsen, J.; Matos, O.; Cesaro, S.; Maschmeyer, G.; Einsele, H.; Donnelly, J.P. ECIL guidelines for the diagnosis of *Pneumocystis jirovecii* pneumonia in patients with haematological malignancies and stem cell transplant recipients. *J. Antimicrob. Chemother.* **2016**, *71*, 2386–2396. [CrossRef] [PubMed]
96. Alanio, A.; Desoubeaux, G.; Sarfati, C.; Hamane, S.; Bergeron, A.; Azoulay, E.; Molina, J.M.; Derouin, F.; Menotti, J. Real-time PCR assay-based strategy for differentiation between active *Pneumocystis jirovecii* pneumonia and colonization in immunocompromised patients. *Clin. Microbiol. Infect.* **2011**, *17*, 1531–1537. [CrossRef]
97. Lu, Y.; Ling, G.; Qiang, C.; Ming, Q.; Wu, C.; Wang, K.; Ying, Z. PCR diagnosis of *Pneumocystis pneumonia*: A bivariate meta-analysis. *J. Clin. Microbiol.* **2011**, *49*, 4361–4363. [CrossRef]
98. Hunter, P.R.; Gaston, M.A. Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. *J. Clin. Microbiol.* **1988**, *26*, 2465–2466. [CrossRef]
99. Le Gal, S.; Damiani, C.; Rouillé, A.; Grall, A.; Tréguer, L.; Virmaux, M.; Moalic, E.; Quinio, D.; Moal, M.-C.; Berthou, C. A cluster of *Pneumocystis* infections among renal transplant recipients: Molecular evidence of colonized patients as potential infectious sources of *Pneumocystis jirovecii*. *Clin. Infect. Dis.* **2012**, *54*, e62–e71. [CrossRef]
100. Helweg-Larsen, J.; Benfield, T.L.; Eugen-Olsen, J.; Lundgren, J.D.; Lundgren, B. Effects of mutations in *Pneumocystis carinii Dihydropteroate synthase* gene on outcome of AIDS-associated *P. carinii* pneumonia. *Lancet* **1999**, *354*, 1347–1351. [CrossRef]
101. Suárez, I.; Roderus, L.; van Gumpel, E.; Jung, N.; Lehmann, C.; Fätkenheuer, G.; Hartmann, P.; Plum, G.; Rybniker, J. Low prevalence of *DHFR* and *DHPS* mutations in *Pneumocystis jirovecii* strains obtained from a German cohort. *Infection* **2017**, *45*, 341–347. [CrossRef]
102. Iliades, P.; Meshnick, S.R.; Macreadie, I.G. Mutations in the *Pneumocystis jirovecii DHPS* gene confer cross-resistance to sulfa drugs. *Antimicrob. Agents Chemother.* **2005**, *49*, 741–748. [CrossRef]
103. Ponce, C.A.; Chabé, M.; George, C.; Cárdenas, A.; Durán, L.; Guerrero, J.; Bustamante, R.; Matos, O.; Huang, L.; Miller, R.F. High prevalence of *Pneumocystis jirovecii Dihydropteroate synthase* gene mutations in patients with a first episode of *Pneumocystis* pneumonia in Santiago, Chile, and clinical response to trimethoprim-sulfamethoxazole therapy. *Antimicrob. Agents Chemother.* **2017**, *61*, e01290-16. [CrossRef]
104. Rabodonirina, M.; Nahimana, A.; Weber, R.; Francioli, P.; Bille, J.; Hauser, P.M. Geographical variation in the prevalence of *Pneumocystis jirovecii Dihydropteroate synthase* mutations within Western Europe. *J. Eukaryot. Microbiol.* **2006**, *53*, S112–S113. [CrossRef]

Pneumocystis jirovecii genetic diversity in a Spanish tertiary hospital

This manuscript was accepted in Medical Mycology on September 9th, 2021. L. Goterris and L. Pasic are joint co-authors and contributed to all aspects of this study (Designing and performing the research, analyzing the data, and preparing the manuscript). M. Guerrero-Murillo, A. Kan, A. Anton, J. Aguilar Company, and I. Ruiz-Camps were involved in the data analysis. M. T. Martin-Gomez and W. Meyer were involved in designing of the research, project administration and funding acquisition. All authors were involved in the revision of the manuscript and support its inclusion in this thesis. The manuscript is added into Chapter 4.2 exactly as it has been submitted to Medical Mycology.

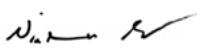
In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lana Pasic

11 October 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Wieland Meyer



11 October 2021

Pneumocystis jirovecii genetic diversity in a Spanish tertiary hospital

Authors: L. Goterris[1-2]*, L. Pasic[3]*, M. Guerrero Murillo[1], Alex Kan[3], A. Anton [1-2], J. Aguilar Company[4-5], I. Ruiz-Camps[4], W. Meyer[3], M.T. Martin-Gomez[1-2#]

*Both authors contributed the same.

Affiliations:

[1]Microbiology Department, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[2]Genetics and Microbiology Department, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.

[3]Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Westmead Hospital, Research and Education Network, Westmead Institute for Medical Research, Westmead, New South Wales, Australia

[4]Infectious Diseases Department, Vall d'Hebron Hospital Universitari, Barcelona, Spain

[5]Oncology Department, Vall d'Hebron Hospital Universitari, Barcelona, Spain

ABSTRACT

Pneumocystis jirovecii is associated with non-noxious colonization or severe pneumonia in immunocompromised hosts. Epidemiological investigations have been hampered by the lack of a standardized typing scheme. Thus, only partial molecular data on Spanish *P. jirovecii* cases are available. Recently a new ISHAM consensus multilocus sequence typing scheme (MLST) targeting *β -TUB, mt26S, CYB,* and *SOD* with a publicly accessible database has been launched to overcome this problem. The molecular epidemiology of *P. jirovecii* from immunocompromised patients either colonized (n=50) or having pneumonia (n=36) seen between 2014 and 2018 at a single center in Barcelona, Spain, was studied. The new ISHAM consensus MSLT scheme was used to investigate the local epidemiology and identify possible unnoticed outbreaks. Mutations in the *DHPS* gene, not included in the scheme but giving information about potential sulpha treatment failure, were also studied. The study assigned 32 sequence types (ST) to 72.2% pneumonia and 56% colonization cases. The most frequent STs were ST21 (18.5%), ST22 (14.8%), and ST37(14.8%). For non- unique STs, ST3, ST30 and ST31 were found only in pneumonia cases, whereas ST27 was associated exclusively to colonization's. Despite 38 patients sharing similar STs, only two were involved in a potential cross transmission event. No *DHPS* mutations were identified. The new consensus typing scheme was useful to ascertain the molecular epidemiology of *P. jirovecii* in our center revealing a high genetic diversity and the potential association of specific STs to colonizationand pneumonia cases.

Lay summary

A newly described MLST scheme aims at providing a standardized tool to study and compare *Pneumocystis jirovecii* epidemiology. A high diversity amongst *P. jirovecii* isolates from patients in Barcelona, Spain, and a potential association between specific STs and infection/colonization were identified.

INTRODUCTION

Pneumocystis jirovecii is an atypical non-culturable fungus taxonomically placed in the subphylum *Taphrinomycotina*, whose only known hosts are humans.[1] *P. jirovecii* is a well-recognized cause of severe pneumonia in an expanding range of immunocompromised hosts (i.e. HIV positive patients with severe lymphopenia, oncological patients, recipients of solid organ transplants, etc.), but it has also been linked to other underlying conditions, and it can be detected in respiratory specimens of asymptomatic individuals regardless of their immune status[2,3]. In any case, the presence of *P. jirovecii* in the respiratory tract of individuals can act as a reservoir, may favor occult person-to- person transmission across time and space, and may act as a potential source of infection for susceptible patients.[4]

Genotyping is considered an important part of epidemiological investigations, as it can give objective information on spatial-temporal transmission between patients, and it is key to characterize isolates that could be associated to specific geographic locations or patient groups.[5] Obtaining a clear map of the *P. jirovecii* molecular epidemiology, however, is extremely difficult, because of several major factors. This fungus cannot be efficiently propagated *in vitro*, so genotyping must be performed directly from clinical samples. Besides this, methodological differences, using typing schemes with a variety of loci, lack of nomenclature standardization of the obtained typing results, and the reportedly unsuccessful amplification of samples with a low fungal content makes it difficult to draw conclusions and to compare data between studies.[6]

Despite having limitations, sequencing of individual nuclear and mitochondrial loci has been used in studies aiming to establish a link between *P. jirovecii* genotypes and clinical traits or geographical locations.[7,8]. Multilocus sequence typing (MLST), with the concatenation of sequences of three or more loci, provides a robust and affordable tool to obtain portable results,with a greater ability to detect genetic diversity.[9]

In the particular case of Spain, little data exists on the molecular epidemiology of *P. jirovecii*. Much of what is known is based on the identification of SNPs in individual genetic loci,[10,11,12,13,14] or on microsatellite analysis,[15,16] a technique that is more simple and less cumbersome than sequencing, but offers results that are difficult to compare between laboratories.[17] Recently, an international consortium has joined efforts to set up a consensus MLST scheme containing the following four genetic loci: *β -TUB, CYB, SOD and mt28S,* in connection with a publicly accessible database aimed to provide a standardized tool for molecular-based epidemiological studies related to *P. jirovecii* that could serve to advance the understanding of the distribution of this elusive microorganism.[6]

Thus, we aimed to contribute to the knowledge of the *P. jirovecii* epidemiology in Spain by studying the genotypes and assigning sequence-types (STs) obtained by using the newly described ISHAM consensus MLST scheme [17] from a cohort of immunocompromised patients attending a single tertiary center in Barcelona. A detailed description of the molecular characteristics of *P. jirovecii* in this cohort is herein provided. In addition, the presence of *DHPS* variants was also investigated, as they may have implications in the therapeutic management of this infection.

MATERIALS AND METHODS

Patients, samples, and laboratory diagnosis

An observational retrospective study was conducted from April 2014 to April 2018. Consecutive immunocompromised patients attending the Hospital Universitari Vall d'Hebron, a 1000-bed tertiary centre (Barcelona, Spain), were included if they had a positive *Pneumocystis* PCR detection in at least one respiratory sample. Patient information and clinical data were registered, including age, sex, underlying condition, baseline immunosuppressive therapy, the use of anti-*Pneumocystis* prophylaxis, *Pneumocystis*-related status (pneumonia or colonization categorized as previously described),[18] disease severity (ICU admission, oxygen requirements, intubation, and mortality) and

healthcare relatedness (see supplementary Table S1). A possible healthcare-related acquisition of *P. jirovecii* was considered if a patient had experienced a >48 h hospital admission within the previous three months.[19] Only one specimen per patient and episode was included (see supplementary Table S1).

Pneumocystis real-time PCR detection was performed on bronchoalveolar lavage (BAL) and oral wash (OW) specimens. DNA was extracted from BAL and OW samples (QIAamp DNA minikit, Qiagen, according to manufacturer's instructions) and *P. jirovecii* DNA was detected using a previously described "in house" real-time PCR assays targeting the nuclear *DHPS* gene.[18,20] Quantitative results were expressed as cycle threshold (Ct). DNA eluates were stored at -20°C until genotyping analysis.

Targeted loci and PCR protocols

The four loci included in the new ISHAM consensus MLST scheme (*β -TUB, CYB, SOD* and *mt28S*) as well as the *DHPS* gene were targeted, following primers and amplification conditions detailed in our previous study.[6,10] Briefly, reactions were carried out in a final volume of 25 µl with 0.4 µM of each primer and 2.5 µl of eluted DNA. DNA was amplified using GoTaq® GreenMaster Mix (Promega Corporation, Madison, WI, USA). Bands of expected size were visualized on2% agarose gels, purified using the BigDye™ Terminator technology according to themanufacturer's instructions (Applied Biosystems) and sequenced. Allele (AT) and sequence types (ST) for each locus sequence were determined based on the public database of the new ISHAM consensus MLST scheme for *P. jirovecii*, [6] available at https://mlst.mycologylab.org, and allele typeidentities with the previously published nomenclature were established.[21,22,23] For each target,sequencing failure was considered if no analyzable sequence was obtained after 2 attempts.

Phylogenetic analysis

Sequences were manually edited and aligned with MEGA v5.2 software (The Biodesign Institute, Tempe, AZ). For each locus, a multiple sequence alignment and a variant calling was performed with the program SNP-sites (https://github.com/sanger-pathogens/snp-sites) in order to identifyvariants, present in our population, and to define genotypes.

Phylogenetic analyses were performed using R software (R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/) and the Phyloviz free online tool (https://online.phyloviz.net/index). Hunter index was calculated for each individual locus and each combination of pure sequence types.[24] Samples presenting ambiguous sequences in more than one locus were excluded from the phylogenetic analysis of concatenated sequences to avoid bias due to uncertain assignment of loci.

Statistical analyses

Continuous variables were expressed as mean and standard deviation (SD). Categorical variables were expressed as proportions (%). Student's t-test and Chi-square or Fisher's tests were used to compare continuous and categorical variables as appropriate respectively. Statistical analysis was performed with Stata statistical package version 15.0 (StataCorp, College Station, Texas, USA).

Ethics

Due to the retrospective non-interventional nature of the study, requirement for informed consent was waived. The study was approved by the Hospital Universitari Vall d'Hebron Ethics Committee (protocol number PR(AG)450/2018).

RESULTS

Eighty-six patients were included in the study, corresponding to 36 (41.9%) pneumonia cases and 50 (58.1%) colonized patients (Table 1). They were obtained from 23 OW and 63 BAL specimens (17 and 19 pneumonia cases, respectively) (Table 1). The cohort was composed predominantly of immunocompromised HIV negative patients (76 out of 86, 88.4%) (Table 1). Demographics and baseline characteristics of patients are listed in Supplementary Table S1.

Samples from the pneumonia cases showed significantly larger fungal loads than those obtained from colonized patients (Ct 32.4 SD 4.0 *versus* Ct 35.5 SD 2.7, respectively; $p<0.0001$). Amplification rates and discriminatory power varied between loci, with successful amplification being lower for samples obtained from colonized patients as compared to those obtained from pneumonia cases (Table 2).

General amplification rates of *β-TUB* and *SOD* were above 70%; three variants were found for each of them, with more than 50% of the sequences clustered in only one variant. A new *β -TUB* variant involving a G to A change in position 87 was detected in one patient sampled in 2017, namely patient HVH46. Regarding mitochondrial loci, high amplification success rates were achieved, reaching 93% for *CYB*, and 95.3% by using new *mt26S* protocol. Compared to *CYB*,allele variants were distributed more evenly across *mt26S* haplotypes, although *CYB* displayed a larger number of allelic types. *DHPS* amplification succeeded in 88.9% of the infection cases and 78% of the colonization cases. No variation in *DHPS* sequences of all studied cases was foundwhen compared to the wild-type sequence (Table 2).

More than one variant was found in at least 1 locus in 30 patients (34.9%). No statistical significance was observed among Ct values of diagnosis via PCR and the presence of sequence mixtures ($p > 0.09$). The locus showing most of the mixed sequences was *mt26S*. Mixed types occurred at the expenses of a few combinations, mainly *mt26S* variant AT3 with variants AT2 or

AT5, combinations of *CYB* variants AT1 and AT6 (85.7%), and combinations of *SOD* AT3 with AT1 (80%). Two patients showed mixed sequences in two loci involving the combination of *CYB* AT1/AT6 with *mt26S* AT3/AT5 or *SOD* AT1/AT3 (see Supplementary Table S1).

When the presence of only one variant per locus was considered for MLST assignment, an ST could be successfully assigned to 18 out of the 36 samples from pneumonia cases (50%), and to 18 out of the 50 samples from colonized patients (36%). The rate of successful ST assignment was similar between the two groups (p=0.268). Overall, when the new ISHAM consensus MLSTscheme was applied to the 36 samples containing only one ST, it achieved a H index of 0.975, and 23 different STs were identified. The most frequently single STs found were ST21, ST22, and ST38(3 patients each).

When mixed sequences in only one locus were also included in the analysis, global success rates increased to 72.2% (26 out of 36) and 56% (28 out of 50) of pneumonia and colonized cases, respectively. Mixed STs were found in 18 out of the 54 samples with successful ST assignment (33.3%), with ST21/ST37 (six samples; 33.3%) and ST22/ST32 (four samples; 22.2%) being the most frequent mixtures (see Supplementary Table S1). Nine out of the 15 (60%) STs found in samples containing more than one genotype in one locus were identified exclusively in mixturesand were detected in only one patient.

Overall, 32 different STs were found alone or in mixes. Among the 54 samples containing at least one ST, the most frequent STs found were ST21 (10 samples; 18.2%), ST22 and ST37 (eight samples -14.5% each), ST32 (seven samples; 12.7%), and ST13 (four samples; 7.3%). Other STs, such as ST3, ST27, ST30, ST31, and ST 38 were found in no more than three patients.

There was no significant association between any particular SNP or concatenated sequence type to specific patient or clinical variables, including: age, gender, baseline condition, pharmacological therapy, use of *Pneumocystis*-active prophylaxis, exposure to the healthcare system or any indicator

of disease severity. Although most STs were homogeneously distributed between colonization and pneumonia cases, for STs found in more than one sample, ST3, ST30 and ST31 were found only in samples from pneumonia cases, whereas ST27 was found only in cases of colonization. Three out of the four cases in which ST13 was identified corresponded to pneumonia cases, and six out of the eight cases in which ST22 was involved corresponded to colonization. Of interest, ST21 was only involved in pneumonia cases when associated to another ST, mainly ST37 (see Supplementary Table S1).

A total of 38 patients presented samples revealing similar STs, alone or in combination (see Supplementary Table S1). Of them, cases closely related in time (at least two of the cases occurring within a period of three months) triggered the investigation of a possible cross-transmission. Up to four potential clusters involving two to eight patients each were identified: two cases in July 2017 sharing ST13, seven cases between June 2014 and January 2015 showing ST21 (in five patients mixed with ST37), six ST22 cases between November 2014 and February 2015 (four mixed with ST32), and two more ST22 cases in February-April 2017. The investigation of a spatial relationshipbetween patients rendered negative results for all but two patients, hospitalized at the same time at the ICU (namely patients HVH21 and HVH22) which shared ST27, rising a reasonable suspicion about a potential cross-transmission. Epidemiological data, however, did not point towards the possibility of a nosocomial outbreak during the study period related to the other cases.

Regarding *DHPS*, amplification succeeded in 71 out of the 86 samples (82.6%). All the sequences identified matched the wild type (Table 2).

DISCUSION

It has long been suggested that there is a need for a consensus method that allows for the reliable differentiation of *P. jirovecii* strains in order to advance our understanding of its epidemiology and

elucidate the long standing question of its mode of transmission.[25] Despite new techniques, such as next generation sequencing, can generate accurate and highly informative molecular data,[26,27] the expensive equipment, high degree of expertise needed, and bioinformatics bottleneck, currently make them unfeasible for most clinical and routine diagnostic applications. Therefore, it is important to apply a methodology that combines appropriate discriminatory power, practicality, and reproducibility between laboratories. Thus, the MLST technique and its new ISHAM MLST scheme for *P. jirovecii* at the current time, presents the most simple and affordable way to achieve this goal.

Little is known of the epidemiology of *P. jirovecii* in Spain. No attempts to apply MLST to ascertain the diversity of this fungus in the country have been communicated to date, and only two studies aiming at genotyping by a combination of microsatellite markers have been performed.[15,16] Moreover, the vast majority of the information from Spain derives from the study of the genetic diversity of individual loci, with the most recent study reporting results from samples collected in Seville between 2006 and 2014.[14] As a result, an updated look on the specific genotypes currently present in Spain is necessary. In addition, the implementation of a consensus genotyping scheme to allow for the comparison of isolates within different regions of Spain is paramount in order to define the current epidemiology and trends of *P. jirovecii* in the country, and subsequently compare its status on a global scale.

The allelic diversity of *SOD* has been poorly studied in Spain, with only one dedicated report published in 2019.[14] Our study confirmed the previously described low genetic diversity of *SOD* genotypes circulating in the country.[14] Despite this, the allele type AT3 was found for the first time in Spain, mainly mixed with *SOD* AT1. This *de novo* presence of the *SOD* AT3 allele may be interpreted as a recent introduction in the area and/or as a variant present exclusively in North- Eastern Spain. As it was mainly present in mixed samples, it is also possible that it might have gone

undetected when other amplification protocols were used. Larger studies might uncover even a greater diversity in the Iberian Peninsula. As described in former reports, *SOD* AT1 was the dominant allele, present in 60.7% of successfully sequenced samples, but in contrast to the findings of Morilla *et al.*,[14] *SOD* AT2 was herein not found exclusively in pneumonia cases.

Of the two nuclear loci studied, the *β -TUB* genotype variation and distribution has not been described previously in Spain. According to former data, the study cohort was dominated by alleles AT2 (61%) and AT1 (37.7%).[23,28,29] Nevertheless, the proportion was inverted as compared to the results obtained from Portuguese HIV positive patients,[23] suggesting some degree of local variation. Even though *β -TUB* has been explored in more than 50 studies, only 5 alleles have been described.[23] In this context, our finding of a new variant AT5 derived from allele AT3, that includesa guanidine to adenine change in position 87, is remarkable. As this locus was studied herein for thefirst time in Spain, it is unknown whether this new variant has emerged *de novo* in recent years or ifit has been endemic to the region for a long time.

Data obtained between 2001 and 2004 from colonization and pneumonia cases diagnosed in patients presenting a variety of underlying conditions in Lisbon and Seville indicated that the *mt28S* AT3 was the most frequent allele type, followed by AT4 and AT2, whereas AT5 was only marginally present.[10,12] This distribution was similar to findings in the United States and the United Kingdom in the same period.[7,30] As derived from our data, predominance still holds true for AT4 and AT2. In contrast, the traditionally infrequent AT5 was present in approximately a fifth of the herein studied samples. It may have experienced an expansion at the expense of an AT3 contraction, which, in the past, has been found to be associated with Spanish HIV positive patients presenting with pneumonia.[10] Although AT5 was found to be more prevalent among colonized patients from Britain,[30] this allele was evenly distributed in samples from colonization and pneumonia cases in the herein studied cohort. As no contemporary data is available, it remains to be

ascertained whether the distribution of *mt28S* allele types found in our study is a hallmark of the local epidemiology,[7] can be attributed to a displacement of allele types that may have occurred in our area over the last two decades, or if it is the result of the near absence of HIV positive patients included in our study, leading to the reduced presence of AT3.

CYB is the target of atovaquone, a second line drug for the prophylaxis and treatment of *P. jirovecii* infections,[31] so mutations in this target may have therapeutic consequences for a subset of patients not candidate to receive other treatments. Despite exhibiting a lower variability than *mt28S* in our cohort, *CYB* proved to be the most polymorphic of all the loci included in the scheme, with up to 7 alleles found distributed in proportions that overlap the ones reported by Sokulska *et al.* in Poland.[32] In agreement with previous reports, *CYB* AT1 was the most prevalent allele,[23,32,33] which could indicate that this allele is, in fact, universally dominant with very few exceptions.[34] Among all the known alleles, the distribution of *CYB* AT8, described for the first time by Maitte *et al.* in 2013,[22] seem to show large geographical variations. It was present in approximately 20% of the samples from French reports,[22,35] but in none of the 31 samples successfully genotyped in a Turkish center.[36] Our study revealed for the first time that this allele is present in Spain, and that it is the second most frequent found for this locus. This finding, however, should be confirmed by studying samples from other regions of Spain.

Negative outcomes and high fungal burdens have been associated with specific SNPs, such as *mt85A, mt85T, SOD110C* and *SOD215C*.[8,37] Although the baseline characteristics were not comparable with the ones reported in a 2010 Portuguese study led by Esteves[8], the SNP distributions of the loci *mt28S, CYB,* and *SOD* in the herein studied cohort were not significantly different from what the authors described, except for the polymorphisms *mt85C* (p= 0.031) and *mt248C* (p=0.0001). Nevertheless, in the present cohort no significant association was found between any SNP or allele type with clinical traits, underlying conditions, severity, outcomes or

fungal load present in the samples. No clear reason exists for such discrepancy; the herein studied group of patients was dominated by HIV-negative individuals who, besides developing a *P. jirovecii* pneumonia with a lower infectious inoculum than HIV positive patients, could be susceptible to infection caused by strains with different degrees of virulence.

The MLST scheme used herein revealed a high heterogeneity in the sequence-types circulating in Barcelona, being able to identify 23 different individual STs, and a further nine STs found only in mixtures. In general, pneumonia and colonization cases were evenly distributed across the 32 STs and showed no statistically significant trend to be clustered in any particular group. Moreover, no relationship was found between STs to any clinical or demographical traits norbetween STs and disease severity. Despite being unable to infer a clear relationship between certain SNPs or STs to potentially greater virulence in this study, specific STs were found to be mostly associated to cases of infection (ST3, ST13, ST21/37, ST30, and ST31) or to colonization (ST22and ST27). As the small numbers prevented us from performing a more robust statistical analysis, further studies at a larger scale are warranted to confirm these findings. It has been suggested that microsatellite length polymorphism would serve better for the purpose of identifying strains with biological traits that make them more infectious.[17]

The attempt to intertwine MLST and clinical data only revealed one potential case of cross-transmission between two colonized patients; this cluster has been described[6] and was not clinically suspected. Other "pseudoclusters" were not supported by epidemiological data, as they were unrelated in time and space. In most of them, *Pneumocystis* was an unexpected finding in outpatients presenting clinical signs and symptoms of respiratory insufficiency and undergoing a bronchoscopy procedure during the routine diagnostic workup of an unrelated baseline pathology. One reason for finding such "pseudoclusters" is that they involved STs that could had been circulating more frequently in the Barcelona area in that moment, but this hypothesis should be

confirmed with more extensive multicenter studies. Another potential explanation is that many of these patients had frequent contact with the healthcare system so, despite no link between patients being identified, we cannot firmly exclude the possibility of a hidden chain of transmission involving asymptomatic carriers. Sanger sequencing is well known to have a limited ability for the detection of mixed genotypes, so minor variants present in the studied samples could have been missed. Moreover, as the analysis was performed retrospectively, a more in-depth search could not be done. To ascertain whether potential case clusters, identified with the help of the MLST analysis,are truly related, next generation sequencing has been used as an additional tool to increase the discriminatory power of MLST; [27] unfortunately, this was not possible in the herein presentedstudy. No environmental reservoir has been identified for this fungus, and airborne person-to-person transmission is thought to be the main route of transmission.[4] Inter-person transmission has been documented for up to 32 months in single centers for given genotypes,[38] and it is assumed tobe involved in the growing number of nosocomial pneumocystis pneumonia outbreaks described in the literature.[39,40] In any case, a periodic genotypic surveillance would be advisable in centers managing high risk patients in order to rule out undetected cross-transmission, and to implement adequate control measures.[39,41]

Given the potential therapeutic implication of mutations that may interfere with the activity of anti-pneumocystis drugs, alterations in the wild-type sequence of the *DHPS* gene have been the focus of a number of studies. Studies carried out between 1990 and 2010, focusing mainly on HIV positive patients, have found a correlation between mutations in *DHPS* gene and treatment failure.[9,42,43] However, contradictory results have arisen from some more recent works including non-HIV patients.[44] As it is unclear whether non-synonymous mutations hamper the *in vivo* effectof co-trimoxazole on *P. jirovecii* infections, further surveillance is warranted. *DHPS* mutations havebeen reported in 3.7-28% of Spanish patients, mainly HIV positive, between 2000 and 2005 in studies using different detection methodologies,[45,46] with a higher proportion of non-wild type

DHPS carriers amongst patients with pneumonia in comparison to those with only colonization (40% vs 22% p=0.112). A lack of association with prior sulpha exposure was also noted, but cases showed a geographical trend, with all *DHPS* mutants found in a Spanish multicenter study two decades ago, in fact, arising from the North Eastern area of Spain.[45] To the best of our knowledge, there is no recent published data on the presence of *DHPS* mutations in *P. jirovecii* positive Spanishsamples, but a declining frequency has been reported in Portugal.[47] The recent widespread use of sulpha prophylaxis in immunocompromised HIV negative patients at risk of developing *Pneumocystis* pneumonia[48,49,50,51,52] warrants surveillance to detect the emergence of mutations that may compromise the efficacy of this drug class.

The following limitations of the present study should be considered. It only included patients from a single Spanish center, and its retrospective design has limited our ability to obtain further clinical details that could have aided in finding links between patient characteristics and genotypes/STs. Furthermore, it also prevented a more in-depth investigation related to potential cross-transmission between patients. Our study, however, is the first to present MLST data of Spanish patients obtained through an easily reproducible, standardized and as such, globally comparable method, opening the door to a more precise knowledge of the molecular epidemiology of this fungus in Spain.

In summary, this work presents the molecular epidemiology of a contemporary cohort of Spanish immunocompromised patients. By using Sanger sequencing, variations in the allele distribution of *mt26S* and *SOD* genetic loci were detected as compared to previous reports, and the types and frequencies of *CYB* and *β -TUB* are presented for the first time from Spanish *P. jirovecii* samples. This study also reveals that *DHPS* mutations may be less prevalent than previously described in Spain. A new MLST consensus scheme was used to characterize the STs circulating among patients attending a single tertiary center in Barcelona, revealing a high genetic diversity.

Further studies are needed to assess the potential association of specific STs with colonization or pneumonia. Nosocomial cross-transmission could have been behind cases sharing the same ST and detected in short periods of time, although the retrospective nature of the work hampered drawing firm conclusions about the origin of these clusters. The use of a standardized MLST scheme supported by a publicly accessible database helps to advance the understanding of theepidemiology, distribution, and dynamics of transmission of this elusive fungus.

ACKNOWLEDGEMENTS:

The authors want to thank Cristina Andres and Carmen Martín for their excellent technical support.

DECLARATION OF INTEREST:

The authors report no conflicts of interest associated with the manuscript. The authors alone are responsible for the content and the writing of the paper.

REFERENCES

1. Ma L, Cissé OH, Kovacs JA. *A Molecular Window into the biology and epidemiology of Pneumocystis spp*. Clin Microbiol Rev. 2018; 31: e00009-18.

2. Morris A *et al. Association of chronic obstructive pulmonary disease severity and Pneumocystis colonization* Am J Respir Crit Care Med. 2004; 170:408-413.

3. Alanio A, Bretagne S. *Pneumocystis jirovecii detection in asymptomatic patients: what does its natural history tell us?* F1000Res. 2017; 6: 739.

4. Choukri F *et al. Quantification and spread of Pneumocystis jirovecii in the surrounding air of patients with Pneumocystis pneumonia.* Clin Infect Dis. 2010; 51: 259-265.

5. Alanio A *et al*. Investigating clinical issues by genotyping of medically important fungi: why and how? *Clin Microbiol Rev*. 2017; 30: 671-707.

6. Pasic L *et al.* Consensus multilocus sequence typing scheme for Pneumocystis jirovecii. *J Fungi (Basel)*. 2020; 6: 259.

7. Beard CB *et al.* Genetic variation in Pneumocystis carinii isolates from different geographic regions: implications for transmission. *Emerg Infect Dis.* 2000; 6:265-272.

8. Esteves F *et al. Identification of relevant single-nucleotide polymorphisms in Pneumocystis jirovecii: relationship with clinical data.* Clin Microbiol Infect. 2010; 16: 878-884.

9. Matos O, Esteves F. *Pneumocystis jirovecii multilocus gene sequencing: findings and implications.* Future Microbiol. 2010; 5: 1257-1267.

10. Montes-Cano MA *et al. Pneumocystis jiroveci genotypes in the Spanish population.* Clin Infect Dis. 2004; 39:123-128.

11. Medrano FJ. *Pneumocystis jirovecii in general population.* Emerg Infect Dis. 2005; 11: 245-250.

12. Esteves F *et al Pneumocystis jirovecii multilocus genotyping pneumonia in patients from Portugal and Spain*. Clin Microbiol Infect. 2008; 14: 356-362.

13. de Armas Y *et al. Low pneumonia diversity of Pneumocystis jirovecii among Cuban population based on two-locus mitochondrial typing.* Med Mycol. 2012; 50: 417-420.

14. Morilla R *et al. Genetic polymorphisms of superoxide dismutase locus of Pneumocystis jirovecii in Spanish Population*. Front Public Health. 2019; 7:292.

15. Parobek CM *et al. Multilocus microsatellite genotyping array for investigation of genetic epidemiology of Pneumocystis jirovecii.* J Clin Microbiol. 2014; 52: 1391-1399.

16. Alanio A *et al. Diversity of Pneumocystis jirovecii across Europe: a multicentre observational study*. EBioMedicine. 2017; 22: 155-163.

17. Gits-Muselli M *et al. Comparison of multilocus sequence typing (MLST) and microsatellite length polymorphism (MLP) for Pneumocystis jirovecii genotyping.* Comput StructBiotechnol J. 2020; 18:2890-2896.

18. Goterris L *et al. Molecular diagnosis of Pneumocystis jirovecii pneumonia by use of oral wash samples in immunocompromised patients: usefulness and importance of the DNA target.* J Clin Microbiol. 2019; 57: e01287-19.

19. Depypere M, Saegeman V, Lagrou K. *Typing of Pneumocystis jirovecii by multilocus sequencing: evidence of outbreak?* Eur J Clin Microbiol Infect Dis. 2016; 35: 911-916.

20. Alvarez-Martínez MJ *et al. Sensitivity and specificity of nested and real-time PCR for the detection of Pneumocystis jiroveci in clinical specimens*. Diagn Microbiol Infect Dis. 2006; 56:153-160.

21. Lee CH, Helweg-Larsen J, Tang X. *Update on Pneumocystis carinii f. sp. Hominis typing based on nucleotide sequence variations in internal transcribed spacer regions of rRNA genes* J Clin Microbiol. 1998; 36: 734-741.

22. Maitte C *et al*. *Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* J Clin Microbiol. 2013; 51: 2843-2849.

23. Esteves F *et al. Population structure of Pneumocystis jirovecii isolated from immunodeficiency virus-positive patients*. Infect Genet Evol. 2010; 10: 192-199.

24. Hunter PR. *et al*. *Reproducibility and indices of discriminatory power of microbial typing methods. J* Clin Microbiol. 1990; 28: 1903-1905.

25. Smulian AG *et al*. *Genetic and antigenic variation in Pneumocystis carinii organisms: tools for examining the epidemiology and pathogenesis of infection.* J Lab Clin Med. 1997; 130: 461-468.

26. Davey JW *et al*. *Genome-wide genetic marker discovery and genotyping using next-generation sequencing.* Nat Rev Genet. 2011; 12: 499-510.

27. Charpentier E *et al. Added value of next generation sequencing for multilocus sequence typing analysis of a Pneumocystis jirovecii pneumonia Outbreak.* Emerg Infect Dis. 2017; 23: 1237-1245.

28. Volpe G *et al. Genetic diversity of Pneumocystis carinii isolated from human immunodeficiency virus-positive patients in Turin, Italy.* J Clin Microbiol. 2001; 39: 2995-2998.

29. Monroy-Vaca EX *et al. Genetic diversity of Pneumocystis jirovecii in colonized Cuban infants and toddlers.* Infect Genet Evol. 2014; 22: 60-66.

30. Miller RF *et al Genotypic variation in Pneumocystis jirovecii isolates in Britain.* Thorax. 2005; 60: 679-682.

31. Maertens J *et al. ECIL guidelines for preventing Pneumocystis jirovecii pneumonia inpatients with haematological malignancies and stem cell transplant recipients.* J Antimicrob Chemother. 2016; 71:2397-2404.

32. Sokulska M *et al. Genotyping of Pneumocystis jirovecii in colonized patients with various pulmonary diseases.* Med Mycol. 2018; 56: 809-815.

33. Nevez G *et al. Pneumocystis primary infection in infancy: additional French data and review of the literature*. Med Mycol. 2020 Feb 1;58(2):163-171. doi: 10.1093/mmy/myz040

34. Szydłowicz M *et al. Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients-preliminary report.* Parasitol Res. 2019; 118: 181-189.

35. Desoubeaux G *et al. Epidemiological outbreaks of Pneumocystis jirovecii pneumonia are not limited to kidney transplant recipients: genotyping confirms common source of transmission in a liver transplantation unit.* J Clin Microbiol. 2016; 54: 1314–1320.

36. Sürgeç E *et al. Genotyping of Pneumocystis jirovecii isolates obtained from clinical samples by multilocus sequencing: a molecular epidemiology study conducted in Turkey*. Arch Microbiol. 2020; 202: 1647-1652.

37. Singh Y *et al. Circulating genotypes of Pneumocystis jirovecii and its clinical correlation in patients from a single tertiary center in India*. Eur J Clin Microbiol Infect Dis. 2017; 36: 1635-1641.

38. Gits-Muselli M *et al. New short tandem repeat-based molecular typing method for Pneumocystis jirovecii reveals intrahospital transmission between patients from different wards*. PloS One. 2015; 10: e0125763.

39. Yiannakis EP, Boswell TC. *Systematic review of outbreaks of Pneumocystis jirovecii pneumonia: evidence that P. jirovecii is a transmissible organism and the implications for healthcare infection control*. J Hosp Infect. 2016; 93: 1-8.

40. Dellière S *et al. Outbreak-causing fungi: Pneumocystis jirovecii*. Mycopathologia. 2020; 185: 783-800.

41. de Boer MGJ, Walzer PD, Mori S. *Healthcare related transmission of Pneumocystis pneumonia: From key insights toward comprehensive prevention.* Transpl Infect Dis. 2018; 20: e12942.

42. Mei Q *et al Failure of co-trimoxazole in Pneumocystis carinii infection and mutations in dihydropteroate synthase gene* Lancet. 1998; 351: 1631-1632.

43. Helweg-Larsen J *et al Effects of mutations in Pneumocystis carinii dihydropteroate synthase gene on outcome of AIDS-associated P. carinii pneumonia.* Lancet. 1999; 354: 1347-1351.

44. Suárez *et al. Low prevalence of DHFR and DHPS mutations in Pneumocystis jirovecii strains obtained from a German cohort.* Infection. 2017; 45: 341-347.

45. Alvarez-Martínez MJ *et al. Pneumocystis jirovecii pneumonia in Spanish HIV-infected patients in the combined antiretroviral therapy era: prevalence of dihydropteroate synthase mutations and prognostic factors of mortality.* Diagn Microbiol Infect Dis. 2008; 62: 34-43.

46. Friaza V *et al*. *Pneumocystis jiroveci dihydropteroate synthase gene mutations among colonized individuals and Pneumocystis pneumonia patients from Spain*. Postgrad Med. 2010; 122: 24-28.

47. Esteves F *et al. Multicentre study highlighting clinical relevance of new high-throughput methodologies in molecular epidemiology of Pneumocystis jirovecii pneumonia.* Clin Microbiol Infect. 2016; 22: 566.e9-566.e19.

48. Cooley L *et al. Consensus guidelines for diagnosis, prophylaxis and management of Pneumocystis jirovecii pneumonia in patients with haematological and solid malignancies, 2014.* Intern Med J. 2014; 44: 1350-1363.

49. Liebling M, Rubio E, Ie S. *Prophylaxis for Pneumocystis jiroveci pneumonia: is it anecessity in pulmonary patients on high-dose, chronic corticosteroid therapy without AIDS?*Expert Rev Respir Med. 2015; 9: 171-181.

50. Brakemeier S *et al.*. P*rophylaxis and treatment of Pneumocystis jirovecii pneumonia after solid organ transplantation.* Pharmacol Res. 2018; 134: 61-67.

51. Braga BP, Prieto-González S, Hernández-Rodríguez J. *Pneumocystis jirovecii pneumonia prophylaxis in immunocompromised patients with systemic autoimmune diseases*. Med Clin (Barc). 2019; 152: 502-507.

52. Prasad GVR *et al. Safety and efficacy of prophylaxis for Pneumocystis jirovecii pneumonia involving trimethoprim-sulfamethoxazole dose reduction in kidney transplantation.* BMC Infect Dis. 2019; 19: 311.

Table 1. Demographics and baseline conditions of patients diagnosed of *Pneumocystis jirovecii* pneumonia versus colonization.

	Total (n=86)	**PJ pneumonia (n=36)**	**PJ colonization (n=50)**	***p* value**
Mean age. Years (SD)	57.3 (15.9)	58.0 (14.6)	56.8 (16.9)	0.734
Sex. No. patients (%)				0.148
Male	52 (60.5)	25 (69.4)	27 (54.0)	
Female	34 (39.5)	11 (30.6)	23 (46.0)	
Underlying condition. No. patients (%)				0.182
Solid tumor	28 (32.6)	14 (38.9)	14 (28.0)	
Hematological malignancies	19 (22.1)	6 (16.7)	13 (26.0)	
Hematopoietic stem cell transplantation	10 (11.6)	2 (5.5)	8 (16.0)	
HIV	8 (9.3)	6 (16.7)	2 (4.0)	
Solid-organ transplant recipients	6 (7.0)	2 (5.5)	4 (8.0)	
Others	6 (7.0)	1 (2.8)	5 (10.0)	
Autoimmune disease	5 (5.8)	3 (8.3)	2 (4.0)	
HIV and solid tumor	1 (1.15)	1 (2.8)	0	
Solid tumor and solid organ transplant	1 (1.15)	0	1 (2.0)	
Solid tumor and hematological malignancy	1 (1.15)	0	1 (2.0)	
HIV, solid tumor and solid organ transplant	1 (1.15)	1 (2.8)	0	
Pharmacological drugs. No. patients (%)				
Systemic corticosteroid therapy	58 (67.4)	22 (61.1)	36 (72.0)	0.288
Biological therapy	28 (32.6)	10 (27.8)	18 (36.0)	0.422
Chemotherapy	36 (41.9)	18 (50.0)	18 (36.0)	0.194
Immunomodulator/ Immunosuppressor agents	19 (22.1)	7 (19.4)	12 (24.0)	0.615
Health-care related episode[a]	58 (67.4)	20 (55.6)	38 (76.0)	0.046
Previous use of prophylaxis against *P. jirovecii*. No. patients (%)	8 (9.3)	0	8 (16.0)	0.012

Abbreviations: PJ- *Pneumocystis jirovecii*; SD- Standard Deviation; No- Number of; HIV: Human Immunodeficiency Virus positive.

a =The episode occurred during a hospitalization or patient have been hospitalized in the previous three months.

Table 2. Amplification rates obtained, single nucleotide polymorphisms distribution found at the five studied loci and discriminatory power by locus.

Locus	Amplification rate (%)	Allele type	Allele type distribution	Single Nucleotide Polymorphism position	Hunter index
β-TUB	69/86 (80.2)	AT2	42 (61.0)	24A; 282G	0.495
		AT1	26 (37.7)	24A; 282A	
		AT5[a]	1 (1.5)	24A; 87A; 282A	
mt26S	82/86 (95.3)	AT3	12 (14.6)	85C; 248C	0.754
		AT4	20 (24.4)	85A; 248C	
		AT2	17 (20.7)	85T; 248C	
		AT5	15 (18.3)	85C; 248T	
		AT11	-	85A; 248T	
		Mixture[b]:	18 (22.0)		
		AT3/AT5	8 (44.4)		
		AT2/AT3	7 (38.9)		
		AT2/AT4	1 (5.6)		
		AT4/AT11	1 (5.6)		
		AT5/AT11	1 (5.6)		
CYB	80/86 (93.0)	AT1	49 (61.3)		0.535
		AT2	5 (6.3)		
		AT4	3 (3.8)	279C; 348A; 516C; 547C; 566C; 838C	
		AT5	1 (1.3)	279C; 348A; 516C; 547C; 566C; 838T	
		AT6	3 (3.8)	279C; 348A; 516C; 547T; 566C; 838C	
		AT7	6 (7.5)	279T; 348A; 516T; 547C; 566C; 838C	
		AT8	6 (7.5)	279C; 348A; 516T; 547C; 566C; 838C	
		Mixture[c]:	7 (8.8)	279C; 348A; 516C; 547C; 566T; 838C	
		AT1/AT4	1 (14.3)	279T; 348A; 516C; 547C; 566C; 838C	
		AT1/AT6	5 (71.4)		
		AT1AT/2/AT6	1 (14.3)		
SOD	61/86 (70.9)	AT1	33 (54.1)		0.507
		AT2	22 (36.1)		
		AT3	1 (1.6)	110C; 215T	
		Mixture[b]:	5 (8.2)	110T; 215C	
		AT1/AT3	4 (80.0)	110T; 215T	
		AT2/AT3	1 (20.0)		
DHPS		Wild type	71/86 (82.6)	165 (55)/A (Thr); 171 (57)/C (Pro)	0

[a] A new variant derived from AT1.

[b] Only were sequenced the patients with best real time PCR Ct values (mean 33.0, SD 3.2).[c] Patients containing more than one sequence type.

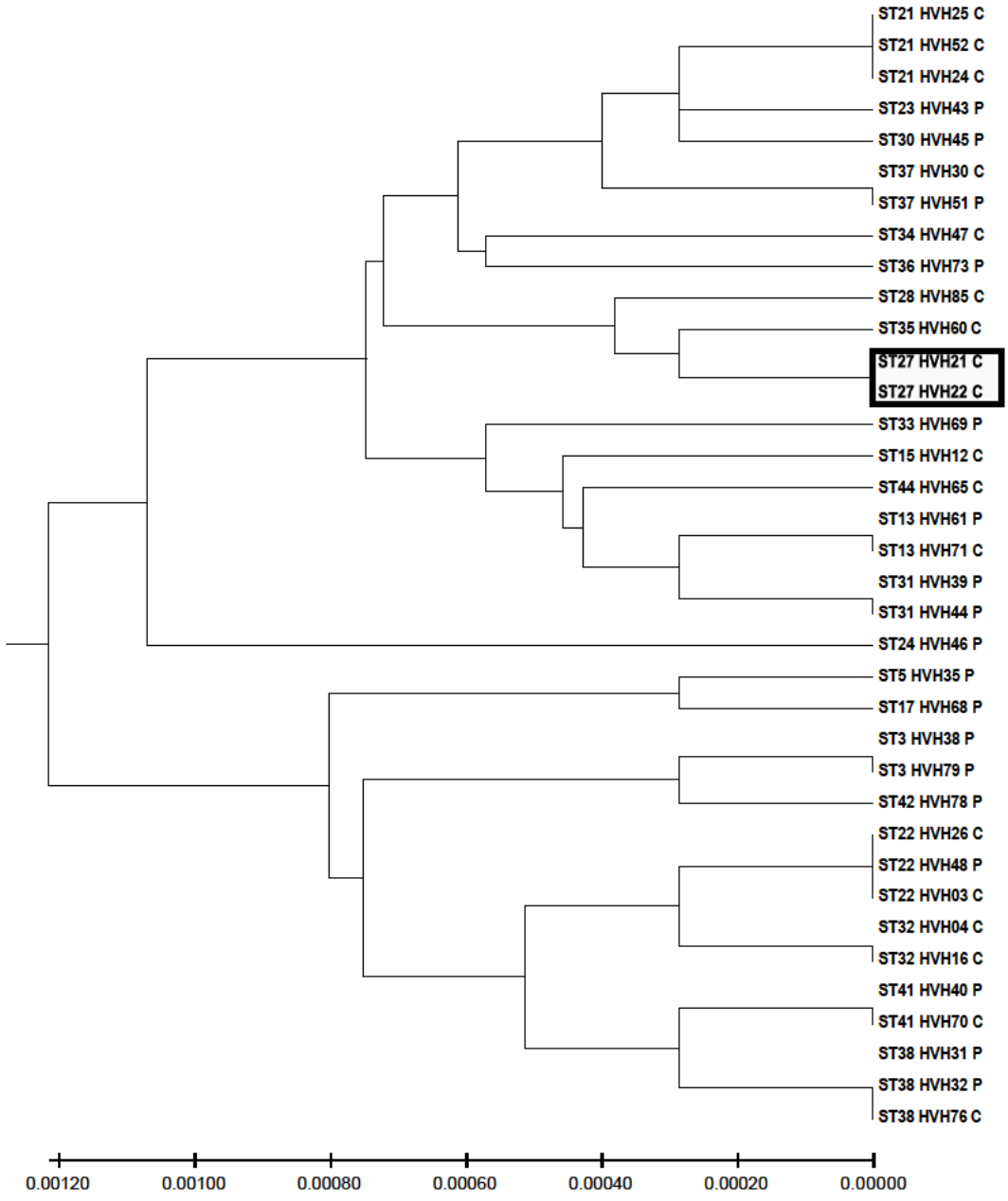


Figure 1. Phylogenetic tree showing genetic distances (Kimura) resulting from the combined analysis of all four MLST loci (*β-TUB, CYB, SOD* and *mt26S*). Only samples harboring one ST are shown. Patients were named consecutively according to the chronology; pneumonia and colonization cases are marked as "P" and "C", respectively. Cases showing molecular and clinical spatial-temporal relationship are highlighted

Supplementary Table S1

PATIENT	Sample Date	mt26S AT	CYB AT	SOD AT	BTub AT	ST ISHAM	Sample	Sex	Age (y)	P jirovecii status	Underlying condition	Previous Hospital Stay	Systemic Steroids	Biologic Agents	Chemotherapy	Immunomodulators	Anti PJ prophylaxis	Anti PJ prophylactic Drug	satO2	Mechanical Ventilation	ICU admission	Exitus
HVH1	Feb 2014	2	1	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	73.6	Colonization	Solid organ malignances Breast cancer	Yes	No	Yes	No	No	No					
HVH2	Mar 2014	3	2	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	28.2	Colonization	Hematopoyetic stem cell transpantation Lymphoma	Yes	No	Yes	No	No	No					
HVH3	Apr 2014	3	1	2	2	22	Bronchoalveolar lavage	Male	60.2	Colonization	Hematological malignances Acute lymphoid leukemia	Yes	Yes	Yes	No	No	Yes	Pentamidina				
HVH4	May 2014	2	1	2	2	32	Bronchoalveolar lavage	Female	75.6	Colonization	Hematological malignances Lymphoma	Yes	Yes	No	Yes	No	No					
HVH5	May 2014	5	1_6	2	No amplification	Not assigned	Bronchoalveolar lavage	Male	62.2	Colonization	Others Pulmonary fibrosis	Yes	Yes	No	No	No	No					
HVH6	Jun 2014	3_5	1	1	2	mix 21/37	Bronchoalveolar lavage	Male	43.6	Pneumonia	HIV debut	No	No	No	No	No	No		98	No	No	No
HVH7	Jun 2014	3_5	1	1	2	mix 21/37	Bronchoalveolar lavage	Male	57.7	Pneumonia	Solid organ malignances Thymus cancer	No	Yes	No	No	Yes	No		89	Yes	Yes	Yes
HVH8	Jun 2014	3_5	1	1	2	mix 21/37	Bronchoalveolar lavage	Male	17.7	Colonization	Hematopoyetic stem cell transpantation Acute myeloid leukemia	Yes	Yes	No	No	Yes	No					
HVH9	Jun 2014	3_5	1	1	2	mix 21/37	Bronchoalveolar lavage	Female	72.0	Colonization	Solid organ malignances Colon cancer	No	No	Yes	No	No	No					
HVH10	Aug 2014	3_5	1	1	2	mix 21/37	Bronchoalveolar lavage	Male	64.8	Colonization	Hematopoyetic stem cell transpantation Myelodisplastic syndrome	No	No	Yes	No	No	No					
HVH11	Sep 2014	5_11	6	3	1	mix 45/46	Bronchoalveolar lavage	Female	44.0	Colonization	Hematological malignances Lymphoma	Yes	Yes	Yes	No	No	No					
HVH12	Oct 2014	4	8	1	1	15	Bronchoalveolar lavage	Male	78.4	Colonization	Hematological malignances Multiple myeloma	Yes	Yes	No	No	Yes	No					
HVH13	Nov 2014	3_2	1	2	2	mix 22/32	Bronchoalveolar lavage	Male	69.1	Colonization	Others Esophagic perforation	No	No	No	No	No	No					
HVH14	Nov 2014	3_2	1	2	2	mix 22/32	Bronchoalveolar lavage	Male	58.5	Colonization	Solid organ malignances Lung cancer	Yes	Yes	No	No	No	No					
HVH15	Nov 2014	4_2	No amplification	2	2	Not assigned	Bronchoalveolar lavage	Female	64.7	Colonization	Solid organ malignances Breast cancer	Yes	No	Yes	No	No	No					
HVH16	Nov 2014	2	1	2	2	32	Bronchoalveolar lavage	Female	59.9	Colonization	Hematological malignances Acute lymphoid leukemia	No	Yes	No	Yes	No	No					
HVH17	Dec 2014	3	1	2_3	2	mix 22/47	Bronchoalveolar lavage	Male	22.1	Colonization	Hematopoyetic stem cell transpantation Acute myeloid leukemia	Yes	Yes	No	Yes	Yes	No					
HVH18	Dec 2014	3_2	1	2	2	mix 22/32	Bronchoalveolar lavage	Female	65.6	Colonization	Hematological malignances Richter Syndrome post chronico lymphoid leukemia	Yes	Yes	Yes	Yes	No	No					
HVH19	Jan 2015	4	7	1_3	2	mix 27/50	Bronchoalveolar lavage	Male	63.6	Colonization	Hematological malignances Acute myeloid leukemia	Yes	No	No	Yes	No	No					
HVH20	Jan 2015	3_2	1	1	No amplification	Not assigned	Bronchoalveolar lavage	Male	63.2	Colonization	Solid organ transplantation Lung transplant	Yes	Yes	No	No	Yes	Yes	Cotrimoxazol				
HVH21	Jan 2015	4	7	1	2	27	Bronchoalveolar lavage	Female	49.2	Colonization	HIV uncontroled disease	No	No	No	No	No	Yes	Cotrimoxazol				
HVH22	Jan 2015	4	7	1	2	27	Bronchoalveolar lavage	Female	20.6	Colonization	Solid organ transplantation Lung transplant	Yes	Yes	No	No	Yes	Yes	Cotrimoxazol				
HVH23	Jan 2015	3_2	1	2	2	mix 22/32	Bronchoalveolar lavage	Male	65.0	Pneumonia	Others Haemophilia A	Yes	Yes	No	No	Yes	No		93	No	No	No
HVH24	Jan 2015	3	1	1	2	21	Bronchoalveolar lavage	Male	53.5	Colonization	Solid organ transplantation Lung transplant	No	Yes	No	No	Yes	Yes	Cotrimoxazol				
HVH25	Jan 2015	3	1	1	2	21	Bronchoalveolar lavage	Female	38.7	Colonization	Autoimmune disease Diffuse scleroderma	No	Yes	No	Yes	Yes	No					
HVH26	Feb 2015	3	1	2	2	22	Bronchoalveolar lavage	Male	18.8	Colonization	Hematopoyetic stem cell transpantation Acute myeloid leukemia	Yes	Yes	Yes	No	No	Yes	Cotrimoxazol				
HVH27	Feb 2015	2	8	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	66.5	Colonization	Solid organ malignances Breast cancer	Yes	Yes	No	Yes	Yes	No					
HVH28	Feb 2015	No amplification	No amplification	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	32.6	Colonization	HIV debut	No	No	No	No	No	No					
HVH29	Sep 2015	4	1	No amplification	1	Not assigned	Bronchoalveolar lavage	Female	54.5	Colonization	Solid organ malignances + Solid	Yes	Yes	Yes	No	Yes	Yes	Cotrimoxazol				

											organ transplantation Lung transplant + squamous carcinoma											
HVH30	Dec 2015	5	1	1	2	37	Bronchoalveolar lavage	Male	65.5	Colonization	So lid organ malignances Lung cancer	No	No	Yes	Yes	No	No					
HVH31	Mar 2016	5	1	2	2	38	Bronchoalveolar lavage	Male	78.5	Pneumonia	So lid organ transplantation Liver transplant	No	No	No	No	Yes	No		80	Yes	Yes	Yes
HVH32	Apr 2016	5	1	2	2	38	Oral Wash	Male	55.1	Pneumonia	So lid organ malignances Lung cancer	Yes	Yes	No	Yes	No	No		90	Yes	Yes	Yes
HVH33	May 2015	3	8	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	71.8	Pneumonia	So lid organ malignances Oesophagus cancer	No	Yes	No	No	No	No		88	Yes	Yes	No
HVH34	May 2015	2	1	No amplification	2	Not assigned	Oral Wash	Male	68.5	Pneumonia	So lid organ malignances Pancreatic cancer	No	No	No	Yes	No	No		96	No	No	No
HVH35	May 2015	2	4	2	2	5	Oral Wash	Male	65.5	Pneumonia	So lid organ malignances Lung cancer	Yes	Yes	Yes	No	No	No		90	No	No	Yes
HVH36	May 2015	3	1	1_3	1	mix 39/40	Bronchoalveolar lavage	Female	58.6	Pneumonia	So lid organ malignances Breast cancer	Yes	Yes	No	Yes	No	No		93	No	No	No
HVH37	Jul 2016	No amplification	No amplification	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	75.4	Colonization	So lid organ malignances choroid melanoma	Yes	Yes	No	Yes	No	Yes	Cotrimoxazol				
HVH38	Aug 2016	4	1	2	1	3	Oral Wash	Male	71.1	Pneumonia	So lid organ malignances Lung cancer	Yes	Yes	No	Yes	No	No		86	No	Yes	Yes
HVH39	Sep 2016	2	1	1	1	31	Bronchoalveolar lavage	Male	51.6	Pneumonia	Hematological malignances Lymphoma	Yes	Yes	Yes	Yes	No	No		89	Yes	Yes	Yes
HVH40	Sep 2016	5	1	2	1	41	Bronchoalveolar lavage	Female	55.6	Pneumonia	Hematological malignances Lymphoma	No	Yes	Yes	Yes	No	No		85	No	No	No
HVH41	Oct 2016	3	8	No amplification	2	Not assigned	Oral Wash	Female	56.1	Colonization	So lid organ malignances Lung cancer	No	Yes	No	Yes	No	No					
HVH42	Nov 2016	3_5	2	1	2	mix 48/49	Bronchoalveolar lavage	Male	51.4	Colonization	Solid organ malignances Astrocytoma	Yes	Yes	Yes	No	No	No					
HVH43	Dec 2016	4	1	1	2	23	Oral Wash	Male	63.4	Pneumonia	Hematopoyetic stem cell transplantation Lymphoma	Yes	Yes	No	No	Yes	No		90	No	No	Yes
HVH44	Jan 2017	2	1	1	1	31	Oral Wash	Male	40.1	Pneumonia	HIV debut	No	No	No	No	No	No		90	No	No	No
HVH45	Jan 2017	2	1	1	2	30	Bronchoalveolar lavage	Male	56.6	Pneumonia	HIV debut	No	No	No	No	No	No		90	No	Yes	No
HVH46	Jan 2017	4	2	1	5	24	Bronchoalveolar lavage	Female	21.3	Pneumonia	Autoimmune disease Ataxia-telangiectasia	No	No	No	No	No	No		91	No	No	No
HVH47	Feb 2017	2	6	1	2	34	Bronchoalveolar lavage	Female	60.6	Colonization	Hematopoyetic stem cell transplantation Lymphoma	Yes	No	Yes	No	No	No					
HVH48	Feb 2017	3	1	2	2	22	Oral Wash	Male	71.2	Pneumonia	Hematological malignances Lymphoma	Yes	Yes	Yes	Yes	No	No		70	No	Yes	No
HVH49	Feb 2017	5	1	No amplification	2	Not assigned	Bronchoalveolar lavage	Female	70.0	Pneumonia	So lid organ malignances Lung cancer	Yes	Yes	Yes	No	No	No		95	No	No	No
HVH50	Mar 2017	3_5	1_6	1	1	Not assigned	Bronchoalveolar lavage	Female	65.3	Pneumonia	So lid organ malignances Breast cancer	Yes	Yes	Yes	Yes	No	No		90	Yes	Yes	No
HVH51	Apr 2017	5	1	1	2	37	Oral Wash	Male	60.9	Pneumonia	HIV + Solid organ malignances + Solid organ transplantation Liver transplant+Lung cancer	Yes	Yes	No	Yes	Yes	No		93	No	No	Yes
HVH52	Apr 2017	3	1	1	2	21	Bronchoalveolar lavage	Male	68.8	Colonization	Hematological malignances Lymphoma	Yes	Yes	Yes	Yes	No	No					
HVH53	May 2017	No amplification	No amplification	No amplification	1	Not assigned	Oral Wash	Male	74.4	Pneumonia	Autoimmune disease Horton arteritis	No	Yes	No	No	No	No		83	No	Yes	Yes
HVH54	Jun 2017	2	2	No amplification	1	Not assigned	Oral Wash	Female	46.1	Pneumonia	HIV + Solid organ malignances uncontroled HIV+Breast cancer	No	No	No	No	No	No		90	Yes	Yes	Yes
HVH55	Jun 2017	3	1_6	1_3	2	Not assigned	Oral Wash	Male	85.0	Colonization	Hematological malignances Myelodispastic syndrome	Yes	Yes	No	No	No	No					
HVH56	Jul 2017	4	1	No amplification	2	Not assigned	Oral Wash	Male	39.8	Pneumonia	So lid organ malignances glioma	Yes	Yes	No	Yes	No	No		84	No	No	No
HVH57	Jul 2017	5	1	No amplification	1	Not assigned	Oral Wash	Male	91.0	Colonization	Hematological malignances Lymphoma	Yes	Yes	No	No	No	No					
HVH58	Jul 2017	5	8	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	40.3	Colonization	Hematopoyetic stem cell transplantation Myelodisplastic	Yes	Yes	No	Yes	No	No					

											syndrome											
HVH59	Jul 2017	2	1	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	60.1	Colonization	Solid organ transplantation Liver transplant	Yes	Yes	Yes	No	Yes	No					
HVH60	Jul 2017	2	7	1	2	35	Bronchoalveolar lavage	Male	53.3	Colonization	Others Leriche syndrome	Yes	Yes	No	No	No	No					
HVH61	Jul 2017	4	1	1	1	13	Bronchoalveolar lavage	Male	37.4	Pneumonia	HIV debut	No	No	No	No	No	No		93	No	No	No
HVH62	Jul 2017	4	1_4	1	1	mix 13/43	Oral Wash	Male	52.9	Pneumonia	Hematological malignances Lymphoma	Yes	Yes	Yes	Yes	No	No		72	No	No	Yes
HVH63	Jul 2017	4	No amplification	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	70.3	Colonization	Hematological malignances Lymphoma	Yes	Yes	No	Yes	No	No					
HVH64	Jul 2017	3_2	1	1	2	mix 21/30	Bronchoalveolar lavage	Male	18.1	Pneumonia	Hematological malignances Acute lymphoid leukemia	No	Yes	No	Yes	No	No		85	No	Yes	No
HVH65	Sept 2017	4	6	1	1	44	Bronchoalveolar lavage	Male	65.7	Colonization	Solid organ malignances Bladder cancer	Yes	Yes	No	No	No	No					
HVH66	Sept 2017	5	1_6	No amplification	1	Not assigned	Bronchoalveolar lavage	Male	42.6	Pneumonia	HIV debut	Yes	No	No	No	No	No		89	No	No	No
HVH67	Sept 2017	3_5	1	1	2	mix 21/37	Oral Wash	Male	64.0	Pneumonia	Solid organ malignances Colon cancer	Yes	No	Yes	Yes	No	No		93	No	No	Yes
HVH68	Nov 2017	2	4	2	1	17	Bronchoalveolar lavage	Female	53.9	Pneumonia	Hematopoyetic stem cell transpantation Myelodisplastic syndrome	No	No	No	No	No	No		91	No	No	No
HVH69	Nov 2017	2	4	1	1	33	Oral Wash	Female	59.5	Pneumonia	Solid organ malignances Breast cancer	Yes	Yes	Yes	Yes	No	No		93	No	No	Yes
HVH70	Nov 2017	5	1	2	1	41	Oral Wash	Female	48.6	Colonization	Solid organ malignances Lung cancer	Yes	Yes	Yes	No	No	No					
HVH71	Dec 2017	4	1	1	1	13	Oral Wash	Male	40.0	Colonization	Solid organ malignances Lung cancer	Yes	Yes	No	Yes	No	No					
HVH72	Dec 2017	3_2	1	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	71.8	Colonization	Hematological malignances Lymphoma	Yes	Yes	Yes	Yes	No	No					
HVH73	Jan 2018	2	8	1	2	36	Oral Wash	Female	65.4	Pneumonia	Hematological malignances Chronic lymphoid leukemia	Yes	No	Yes	Yes	No	No		93	No	No	No
HVH74	Jan 2018	No amplification	No amplification	1_3	No amplification	Not assigned	Bronchoalveolar lavage	Female	62.1	Pneumonia	Autoimmune disease dermatomyositis	No	Yes	No	Yes	Yes	No		93	Yes	Yes	No
HVH75	Feb 2018	2	1_2_6	2	2	Not assigned	Bronchoalveolar lavage	Female	63.6	Colonization	Solid organ malignances Lung cancer	Yes	Yes	No	No	No	No					
HVH76	Feb 2018	5	1	2	2	38	Bronchoalveolar lavage	Male	58.9	Colonization	Others Pulmonary nodule	No	No	No	No	No	No					
HVH77	Mar 2018	4	1	No amplification	1	Not assigned	Bronchoalveolar lavage	Female	69.4	Pneumonia	Solid organ transplantation Liver transplant	Yes	Yes	No	No	Yes	No		98	Yes	Yes	Yes
HVH78	Mar 2018	4	2	2	1	42	Bronchoalveolar lavage	Male	54.2	Pneumonia	Solid organ malignances Lung cancer	Yes	Yes	No	Yes	No	No		75	No	No	Yes
HVH79	Apr 2018	4	1	2	1	3	Oral Wash	Male	82.1	Pneumonia	HIV debut	No	No	No	No	No	No		96	No	No	Yes
HVH80	May 2018	4_11	1	No amplification	No amplification	Not assigned	Oral Wash	Female	57.7	Colonization	Solid organ malignances Breast cancer	Yes	Yes	No	Yes	No	No					
HVH81	Jun 2018	5	1	No amplification	1	Not assigned	Bronchoalveolar lavage	Female	31.7	Colonization	Hematopoyetic stem cell transpantation Lymphoma	Yes	Yes	Yes	No	Yes	No					
HVH82	Jun 2018	4	1_6	1	1	mix 13/44	Oral Wash	Male	74.5	Pneumonia	Solid organ malignances Lung cancer	Yes	No	No	Yes	No	No		90	No	No	Yes
HVH83	Jun 2018	5	7	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	53.4	Colonization	Autoimmune disease Psoriatic arthritis	No	Yes	No	Yes	No	No					
HVH84	Jul 2018	4	5	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Male	69.6	Colonization	Others femoral stent placement	Yes	No	No	No	No	No					
HVH85	Jul 2018	4	7	1	1	28	Bronchoalveolar lavage	Female	57.1	Colonization	Solid organ malignances + Hematological malignances Liver cancer + Acute lymphoid leukemia	Yes	Yes	No	No	Yes	No					
HVH86	Ago 2018	5	1	No amplification	No amplification	Not assigned	Bronchoalveolar lavage	Female	66.5	Colonization	Hematological malignances Acute myeloid leukemia	Yes	No	No	Yes	No	No					

NAm: No Amplification; NAs: Not assigned; OW: Oral Wash; LBA: Bronchoalveolar lavage; M: Male; F: Female; C: Colonization; P: Pneumonia; HSTC: Hematopoyetic stem cell transplantation; ALL: Acute lymphoid leukemia; AML: Acute myeloid leukemia; MDS: Myelodisplastic syndrome; MM: Multiple myeloma; SOT: Solid organ transplantation; CLL: Chronic lymphoid leukemia

Chapter 5

Applying the novel ISHAM consensus MLST scheme to set the basis for a global population genetic analysis of the human pathogenic fungus *Pneumocystis jirovecii*

5.1 Introduction

Pneumocystis jirovecii can cause the debilitating fungal disease Pneumocystis Carinii Pneumonia (PCP), also known as Pneumocystis Jirovecii Pneumonia (PJP), which can be life threatening for immunocompromised individuals [1]. Although once the major cause of mortality in AIDS patients, in western societies, with the introduction of HAART therapy and widespread prophylaxis schemes, mortality rates in the HIV positive population has declined. Nowadays the largest known risk factor for PCP fatalities in acquiring a nosocomial PCP infection by individuals with poor host immunity, such as transplant recipients, autoimmune and cancer patients [2]. To circumvent this, implementation of blanket prophylaxis regimes for high-risk groups in attempts to minimize PCP infections and, in turn, avoid outbreaks in hospitals has been proposed [3, 4], although this is not practical in all circumstances, and also comes with its own implications, such as increasing resistance to trimethoprim sulfamethoxazole (TMZ-SMZ) [5, 6]. Due tothe inability to be cultured, molecular typing has been the key to advancing the understanding of *P. jirovecii* transmission patterns, epidemiological profiles, virulence, and resistance [7]. Agold standard for genotyping across many pathogenic microorganisms has been multilocus sequence typing (MLST), which has helped to identify and characterize genetic profiles of strains [3, 8]. To do so, MLST uses PCR to amplify and subsequently sequence a combination of several housekeeping loci from an organism. It is highly reproducible and affordable [9]. In the case of PCP outbreaks, MLST has enabled to map hospital transmissions and classify index cases, as well as give insights into associations of specific sequence types with clinical features [2, 10, 11].

Over the past two decades of *Pneumocystis* epidemiological studies, molecular advances have enabled to address several public health and clinical questions, including: are some strains causing PCP are more pathogenic or resistant to treatment than others, determine

the risk of colonizing *P. jirovecii* strains to act as a reservoir, and better understand the spread of certain strains [12]. Despite all the advances, and even there is evidence for some strong associations between those factors, some reports remain inconsistent or contradictory, making further large-scale population studies essential.

A huge drawback in epidemiological *Pneumocystis* research was, that there hasn't been a formally accepted consensus typing scheme and no global database, making a general comparison and collaboration between laboratories difficult, as each laboratory did choose the genetic loci they wished to use in an MLST scheme. This fact has only recently been overcome by establishing a standardized MLST scheme, the ISHAM MLST scheme for *P. jirovecii*, and promote it as a global consensus scheme [13]. The consensus scheme utilizes single-round PCR and amplifies four genetic loci, the *β-tubulin* gene (*β- TUB*), *cytochrome b oxidase* gene (*CYB*), large submitochondrial RNA gene (*mt26S*) and *superoxide dismutase* gene (*SOD*). Subsequently, a database has been established, which can be found at https://mlst.mycologylab.org/, storing the sequence date of already observed sequence types. Goterris *et al.* 2021, were the first to publish a population analysis using this scheme and describes the diversity of allele and sequence types from a single Hospital in Barcelona, Spain [14]. The MLST scheme was able to provide insights into the local epidemiology and offer further information into associations between specific genotypes and clinical disease characteristics.

The herein reported study aimed to conduct an epidemiological investigation of all *P. jirovecii* samples for which MLST sequence data have been submitted to the ISHAM MLST database using the consensus MLST *P. jirovecii* scheme. The database encompassed MLST sequence submissions from a variety of samples source sites, from both colonized and

infected individuals, originating from five different countries scattered across three continents. This is adding to our understanding of the global *P. jirovecii* population structure.

5.2 Methods and Materials

5.2.1 Samples studied

This study combined MLST data obtained from *P. jirovecii* positive samples from various patient cohorts (Appendix Table 1). The largest dataset originated form the "Spanish" cohort, which is comprised of 53 samples, 41 bronchoalveolar lavage (BAL) samples and 12 oropharyngeal washes (OW), which had been obtained from the Hospital Universitari Vall d'Hebron, Madrid, between 2014 and 2018, including 25 positive PCP cases and 28 *P. jirovecii* colonisations. This sample set resulted in 72 independent sequences generated when including samples of mixed infections. This dataset was previously reported in Goterris *et al.,*2021 together with in depth sample metadata [14]. The second largest dataset originated fromthe "Chilean" cohort, which is comprised of 41 *P. jirovecii* positive BAL samples, resultingin 44 sequences generated, which had been obtained from patients from hospitals across Santiago de Chile, between 2004 and 2014. These two major datasets were complimented with two minor additional datasets. The first consists of a "Brazilian" cohort, containing 17.
P. jirovecii positive samples, which were collected at the Fundação Oswaldo Cruz (Fiocruz) in Rio de Janeiro from 2016 to 2017, The second dataset was generated from a "Oceania" cohort, including 7 *P. jirovecii* positive Australian samples from two Sydney hospitals,resulting in 9 identified sequences, and 2 *P. jirovecii* positive samples from the AucklandCity Hospital, New Zealand. There were several additional *P. jirovecii* positive samples from which not all four the ISHAM MLST consensus loci could be successfully sequenced, these

samples will not be included in the sequence type analysis, but unique allele types obtained from these samples will be reported.

5.2.2 DNA extraction

For all BAL specimens, the BAL DNA Isolation Kit (cat. No. 46 200, Norgen, Canada) was used for DNA extraction, with the inclusion of an added liquid nitrogen grinding step, as described by Irinyi *et al.*, 2020. IS specimens were extracted using the QIAamp DNA Mini kit (Qiagen) following the manufacturer's *instructions*.

5.2.3 *Pneumocystis jirovecii* confirmation

All samples were diagnostically confirmed to have *P. jirovecii* during the initial hospital diagnostics workup, using inhouse methods. The presence of *P. jirovecii* DNA was confirmed prior conducting any analysis for the current study, using PCR of the *mt26S* gene with primers PneumoLSU_F and PneumoLSU_R [13]. Negative controls using distilled water were included to monitor for cross-contamination during DNA extraction and purification. All samples were processed inside a biosafety cabinet using sterile precautions to avoid contamination.

5.2.4 PCR Amplification

The four genetic loci of the ISHAM MLST *P. jirovecii* consensus scheme, *β-TUB, CYB, mt26S, SOD,* were separately amplified [3]. The amplifications were carried out in volumes

of 25 µl per PCR reaction, using 10 X buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$ and 0.01% *w*/*v* gelatine), 50 nM $MgCl_2$, 2 mM dNTPs, 10 ng/µl of each primer, 5 U/µl BIOTAQ DNA (Bioline) polymerase and 10 µl of genomic DNA. The *β-TUB* locus was amplified with the forward primer PneumoβTub_F, 5´-TCATTAGGTGGTGGAACGGG-3´, and the reverse primer PneumoβTub_R, 5´-ATCACCATATCCTGGATCCG-3´. The *SOD* locus was amplified with the forward primer MnSODFw, 5´-GGGTTTAATTAGTCTTTTTAGGCAC-3´and the reverse primer MnSODRw 5´-CATGTTCCCACGCATCCTAT-3´. Both loci were amplified using the following PCR conditions, an initial denaturing step at 95°C for 3 minutes; followed by 45 cycles: 94°C for 30 seconds, 60°C for 45 second, and 72°C for 45 seconds, followed by a final extension step of 72°C at 7 minutes. The *CYB* locus was amplified with the forward primer CytbFw, 5´-CCCAGAATTCTCGTTTGGTCTATT-3´, and the reverse primer CytbRw 5´-AAGAGGTCTAAAAGCAGAACCTCAA-3. The *mt26S* locus was amplified with the forward primer PneumoLSU_F 5'-TCAGGTCGAACTGGTGTACG-3' and the reverse primer PneumoLSU_R 5'-TGTTCCAAGCCCACTTCTT-3'. Both loci were amplified using the following PCR conditions, an initial denaturing step at 95°C 3 min; followed by 45 cycles: 94°C for 30 seconds, 55°C for 45 seconds, 72°C for 45 seconds; followed by a final extension step of 72°C for 7 minutes. Successfully amplified loci were visualized on 1.5% agarose gel electrophoresis, and molecular sizes were confirmed by the comparison with the 1 Kb Plus DNA Ladder (Invitrogen), which was loaded to each gel as a molecular size standard. The obtained fragments were visualized using UV illumination on a transilluminator.

5.2.5 DNA Sequence analysis

Bidirectional sequencing was performed at Macrogen Inc., Seoul, South Korea. The obtained sequences where then assembled and edited using the software package Sequencher ver. 5.4.6 (Gene Codes Corporation). The cleaned-up sequences for each locus were aligned with the program CLUSTALW part of the software MEGA version 10.1 and compared to reference sequences published in Pasic *et al.*, 2021 [15, 16].

5.3 Results

The herein conducted population genetic analysis included samples from 123 patients combining four patient cohorts from Australia, Brazil Chile, Spain, and New Zealand. The patients were a mixture of both colonised and PCP infected individuals, from which BAL, IS and OW samples had been obtained, collected from 2004 until 2018. The novel ISHAM MLST consensus scheme for *P. jirovecii* identified 48 sequence types amongst the 123 samples, representing a high degree of genetic variability within this population, as indicated by the H-index of 0.9677.

5.3.1 Allele Type Diversity

In terms of allelic diversity, the four cohorts resulted in a variety of allele types, with a total of 34 variants among the four genetic loci studied. The highest diversity was seen in the *CYB* locus, with a total of 12 allele types, followed by the *mt26S* locus with 11 allele types, the *SOD* locus with 6 allele types, and the *β-TUB* locus which showed the least diversity with 5 allele types. The study identified a total of 21 SNPs (Table 5.1). The *CYB* locus harboured

SNP variants only between cytosine and thymine, a trend not seen in any of the other genetic loci. Several country specific alleles were detected in each allelic group, that had not been published before, including AT3 and 4 from locus *β-TUB*, AT12 and AT13 from loci *CYB*, AT9 and AT10 from *mt26S*, well as AT4, AT5 and AT6 from locus *SOD*.

Table 5.1. Table of allele types reported within the ISHAM MLST consensus scheme, with their respectiveSNP locations, rate at which it appears within the samples in the database, and countries they have been detectedin. *Signifies that this allele type was identified, but from an isolate which did not have all four allelessequenced successfully.

	ALLELE TYPES	SNP LOCATIONS						NUMBER (%) OF ISOLATES	COUNTRIES AT DETECTED
β -TUB		87	281	282					
	AT1	G	T	A				66 (45.5%)	AU, BR, CL, ES, NZ
	AT2	G	T	G				74 (51%)	AU, BR, CL, ES
	AT3	G	T	T				3 (2.1%)	BR
	AT4	G	A	T				1 (0.7%)	BR
	AT5	A	T	A				1 (0.7%)	ES
CYB		12	249	280	290	299	571		
	AT1	C	C	C	C	C	C	69 (47.6%)	AU, CL, ES
	AT2	C	C	C	C	C	T	11 (7.6%)	BR, CL, ES
	AT4	C	C	C	T	C	C	15 (10.3%)	AU, BR, CL, ES
	AT5	T	T	C	C	C	C	10 (6.9%)	CL
	AT6	C	T	C	C	C	C	6 (4.1%)	ES
	AT7	C	C	C	C	T	C	6 (4.1%)	ES
	AT8	T	C	C	C	C	C	16 (11%)	BR, CL, ES
	AT9	T	T	C	T	C	C	3 (2.1%)	CL
	AT10	T	C	C	C	C	T	3 (2.1%)	CL
	AT11	C	T	T	C	C	C	*	BR
	AT12	T	T	T	C	T	T	4 (2.8%)	BR
	AT13	C	T	T	C	T	T	3 (2.1%)	BR
mt26S		27	51	87	250	290	291		
	AT1	X	X	C	C	G	A	3 (2.1%)	BR
	AT2	X	X	T	C	A	A	29 (20%)	AU, CL, ES
	AT3	X	X	C	C	A	A	25 (17.2%)	BR, ES
	AT4	X	X	A	C	A	A	55 (37.9%)	AU, CL, ES, NZ
	AT5	X	X	C	T	A	A	21 (14.4%)	BR, CL, ES
	AT6	C	C	T	C	A	A	4 (2.8%)	CL
	AT7	X	X	C	C	A	T	1 (0.7%)	NZ
	AT8	X	X	T	C	A	T	3 (2.1%)	BR
	AT9	X	X	C	G	A	A	*	BR
	AT10	X	X	A	G	A	A	3 (2.1%)	BR
	AT11	X	X	A	T	A	A	1 (0.7%)	ES
SOD		51	52	53	156	326	566		
	AT1	C	C	C	T	G	T	69 (47.6%)	AU, BR, CL, ES, NZ
	AT2	T	C	C	C	G	T	54 (37.2%)	AU, BR, CL, ES, NZ
	AT3	T	C	C	T	G	T	5 (3.5%)	ES
	AT4	C	X	X	T	G	T	*	BR
	AT5	T	C	C	C	G	C	6 (4.1%)	BR
	AT6	C	C	C	T	A	T	11 (7.6%)	BR

Individual genotypic analysis of the four different allele types was undertaken (Figure 5.1). Clusters were evident between all loci, two major branching clusters in the *β-TUB* locus

(Figure 5.1A) and *SOD* locus (Figure 5.1D), while the *CYB* locus (Figure 5.1B) and the *mt26S* locus (Figure 5.1C) showed three distinct clusters.

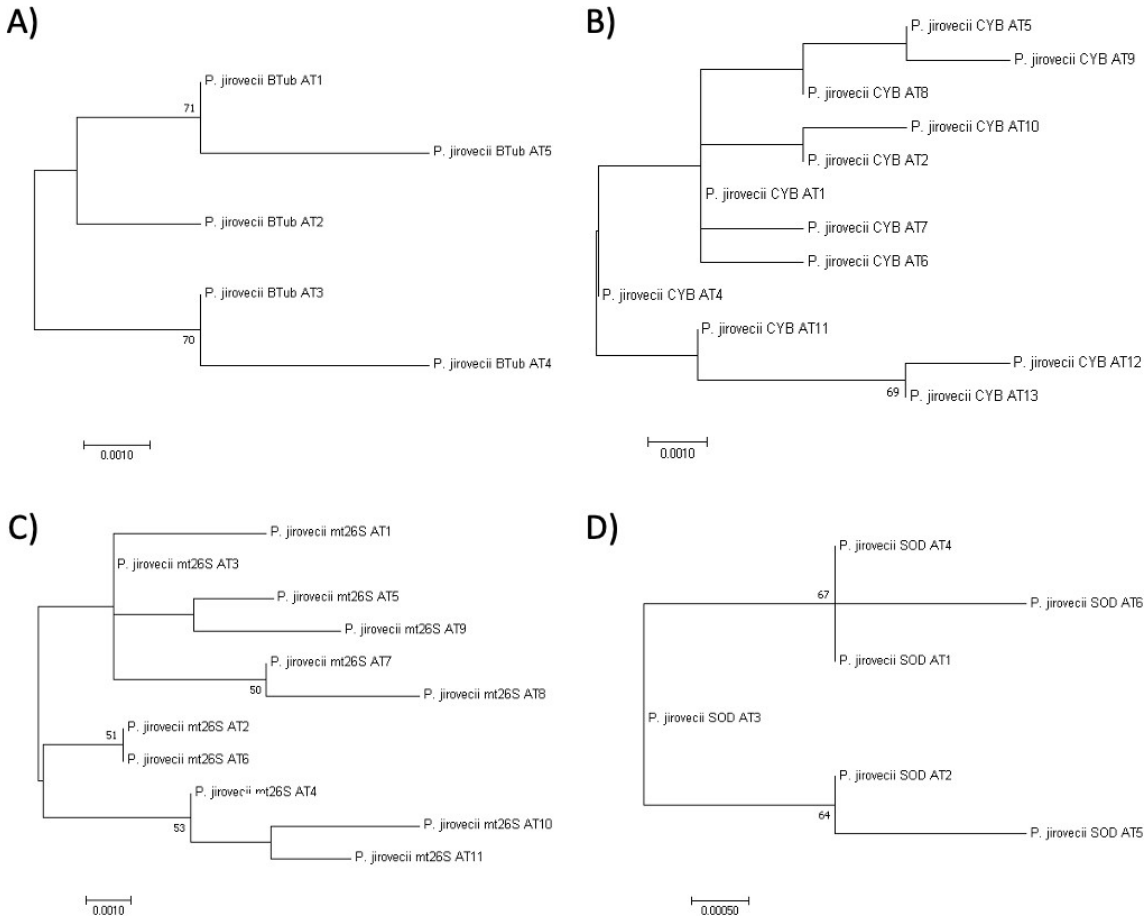


Figure 5.1. Individual dendrogram of the allele types detected on the ISHAM database, of all four loci. A) *β-TUB*, B) *CYB*, C) *mt26S*, and D) *SOD*. Clusters were evident between all loci. The numbers given above the branches are frequencies (>60%) with which a given branch appeared in 1,000 bootstrap replications.

Some specimens had more than one *P. jirovecii* genotype per sample, with only specimens which had one nucleotide change, in only one allele, having been considered. This resulted in 24 samples which had as such a mixed infection with two sequence types. All samples which had mixed infections that included more than one base pair change were not included, as there was no reliable way to accurately assign allele or sequence types. In total, there were 43 samples which were positive for *P. jirovecii*, but unable to have genotyping

analysis being undertaken due to multiple SNP differences at the same position (data not included).

There were several variants detected in each allele type, except for the *β-TUB* locus, with a total of five samples with two allele types within the *CYB* locus, 13 samples with two allele types within the *mt26S* locus, and five samples with two allele types within the *SOD* locus. Table 5.2 shows the combination of mixed infections identified in a single sample.

Table 5.2. Description of which allele types were identified in mixed alleles, and their incidence rate

Locus	**Mixed allele type and combinations and rate (%)**		
CYB	AT1 and AT4 (40%)	AT4 and AT10 (20%)	AT5 and AT8 (40%)
mt26S	AT2 and AT3 (31%)	AT3 and AT5 (61%)	AT5 and AT11 (8%)
SOD	AT1 and 2 (40%)	AT1 and AT3 (40%)	AT2 and AT3 (20%)

High genetic variability within the allele types has led to a combination of 48 sequence type among the population studied, as shown in Table 5.3. There are four sequence types which appear across three of the countries studied, including ST3, ST5, ST13 and ST15;while two sequence types appear across two countries, which are ST2 and ST17. The remainder of the sequence types described only appear in a single country.

Table 5.3. The allele type and respective sequence types detected of all fully sequenced isolates submitted in the ISHAM database.

***β-TUB* AT**	***CYB* AT**	***mt26S* AT**	***SOD* AT**	**Sequence Type**	**Countries* ST Detected**	**Number (%) of isolates**
1	10	2	2	1	CL	1 (0.7%)
2	2	4	5	2	BR, CL	3 (2.1%)
1	1	4	2	3	AU, CL, ES	16 (11%)

1	8	5	1	4	CL	2 (1.4%)
2	4	2	2	5	AU, CL, ES	8 (5.5%)
1	9	5	1	6	CL	3 (2.1%)
2	5	4	1	7	CL	3 (2.1%)
2	8	4	1	8	CL	4 (2.8%)
2	5	2	6	9	CL	3 (2.1%)
4	12	5	6	10	BR	1 (0.7%)
1	13	10	6	11	BR	3 (2.1%)
1	5	6	6	12	CL	4 (2.8%)
1	1	4	1	13	AU, ES, NZ	8 (5.5%)
3	12	8	5	14	BR	3 (2.1%)
1	8	4	1	15	BR, CL, ES	6 (4.1%)
2	2	4	2	16	CL	4 (2.8%)
1	4	2	2	17	CL, ES	2 (1.4%)
1	4	1	1	18	BR	3 (2.1%)
1	8	3	2	19	BR	3 (2.1%)
1	1	7	2	20	NZ	1 (0.7%)
2	1	3	1	21	ES	10 (6.9 %)
2	1	3	2	22	ES	8 (5.5%)
2	1	4	1	23	ES	1 (0.7%)
5	2	4	1	24	ES	1 (0.7%)
2	7	4	1	27	ES	3 (2.1%)
1	7	4	1	28	ES	1 (0.7%)
2	10	2	2	29	CL	1 (0.7%)
2	1	2	1	30	ES	2 (1.4%)
1	1	2	1	31	ES	2 (1.4%)

2	1	2	2	32	ES	6 (4.1%)
1	4	2	1	33	ES	1 (0.7%)
2	6	2	1	34	ES	1 (0.7%)
2	7	2	1	35	ES	1 (0.7%)
2	8	2	1	36	ES	1 (0.7%)
2	1	5	1	37	ES	8 (5.5%)
2	1	5	2	38	ES	3 (2.1%)
1	1	3	1	39	ES	1 (0.7%)
1	1	3	3	40	ES	1 (0.7%)
1	1	5	2	41	ES	2 (1.4%)
1	2	4	2	42	ES	1 (0.7%)
1	4	4	1	43	ES	1 (0.7%)
1	6	4	1	44	ES	2 (1.4%)
1	6	5	3	45	ES	1 (0.7%)
1	6	11	3	46	ES	1 (0.7%)
2	1	3	3	47	ES	1 (0.7%)
2	2	3	1	48	ES	1 (0.7%)
2	2	5	1	49	ES	1 (0.7%)
2	7	4	3	50	ES	1 (0.7%)

*Note: AU=Australia; BR=Brazil; CL= Chile; ES = Spain; NZ=New Zealand

Genotyping analysis of all the samples (Figure 5.2) demonstrated a general trend that sequence types from the same country group together, but also showing a widely spread of the sequence types throughout the tree and with no obvious uniform grouping patternsbetween the clades.

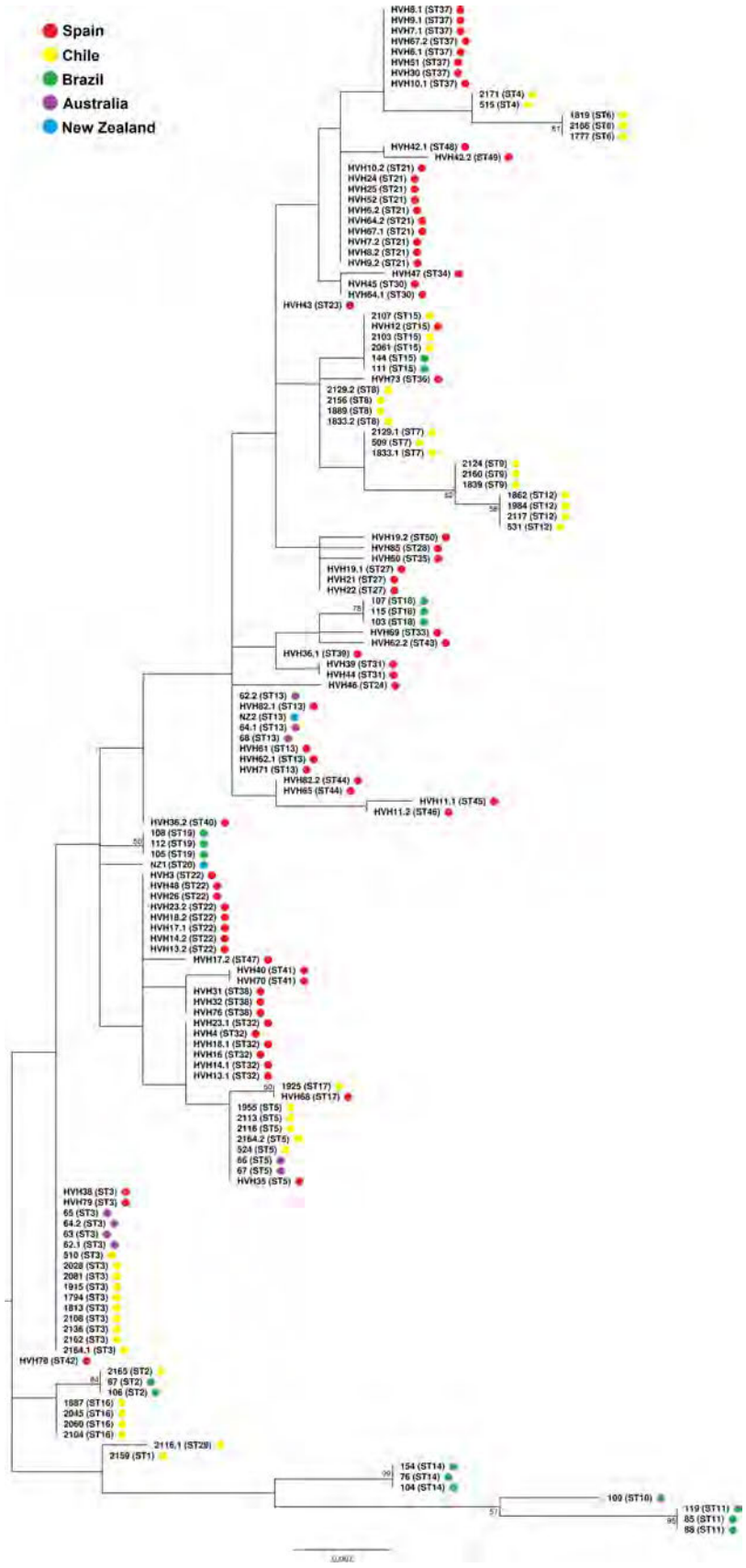


Figure 5.2. A dendrogram of all studied samples for which currently MLST data are deposited in the ISHAM MLST *P. jirovecii* database. The numbers given above the branches are frequencies (>60%) with which a given branch appeared in 1,000 bootstrap replications.

5.4 Discussion

It is more and more becoming evident that genotyping of pathogenic microbes is important for the management of public health responses to emerging disease outbreaks, especially as invasive fungal diseases, such as *P. jirovecii* are posing a major health risk in immunocompromised patients. As nosocomial PCP infections carry high morbidity and mortality rates, molecular microbiology has helped clinicians to better prevent and predict outbreaks. Genotyping of variants has helped develop the understanding of the epidemiology of infectious pathogens, especially establishing the relationships between strains, the spread of disease, and further understand the impact of recombination between strains and their distribution in the general community, and its impact on healthcare overall.

The study of allele and sequence types can also give insights into clinical outcomes, resistance to treatment and effectiveness of prophylaxis in high-risk patient cohorts. To date, extensive research into genotyping has been undertaken for *P. jirovecii,* with almost 70 genotyping publications, describing the use of over 31 different genotyping schemes. The reported data within the literature is at times contradictory, especially in the discussions concerning specific genotypes and their impact on health outcomes. A major hurdle hindering collaborative research and the ability to strengthen the impact of the exiting data was the lack of a globally accepted consensus MLST scheme for *P. jirovecii,* such as the ones existing for *Cryptococcus neoformans* and *C. gattii* [17].

The recent establishment of a MLST consensus scheme and the formation of a specific database hope to alleviate some of these issues and to increase genomic typing, while continuing to be the gold standard. During its development, the scheme was shown to have a high discriminatory power, with a H-index of 0.975 and showed success to obtain genotyping

data of samples from a range of source sites, including OW, IS and BAL. Furthermore, the scheme was found to be successful in sequencing both *P. jirovecii* from infected and colonized hosts. The MLST scheme was optimized to increase amplification rates, to overcome limitations of previous MLST schemes, especially in problematic loci, such as the *SOD*, resulting in high amplification and sequencing rates [13, 18, 19]. Since then, Goterris *etal.*, 2021, used the MLST scheme to assess the diversity of *P. jirovecii* obtained from a major Spanish hospital, detecting 24 individual sequence types [14]. This study set the path for the use of a global MLST allele and sequence type database, to enable comparisons to beextended beyond a single hospital site to expand the current population studies of *P. jirovecii* and its patterns of spread globally. To date, the ISHAM MLST database for *P. jirovecii* has a total of 145 submissions with complete sequencing for all four loci, covering five countries scattered across three continents. Therefore, a global comparison of all complete submissions was undertaken herein. To the best of our knowledge, this is the most geographically widespread analysis completed by using the same typing scheme.

The Spanish isolates are extensively covered as an individual cohort in the Goterris *et al.* study, and therefore, the Spanish genotypes will only be described in the global context herein [14]. The allele types reported globally in the current study expanded on those being endemic in Spain, with a high heterogeneity noted within the studied population and new alleles described for all four genetic loci studied (Figure 5.3).

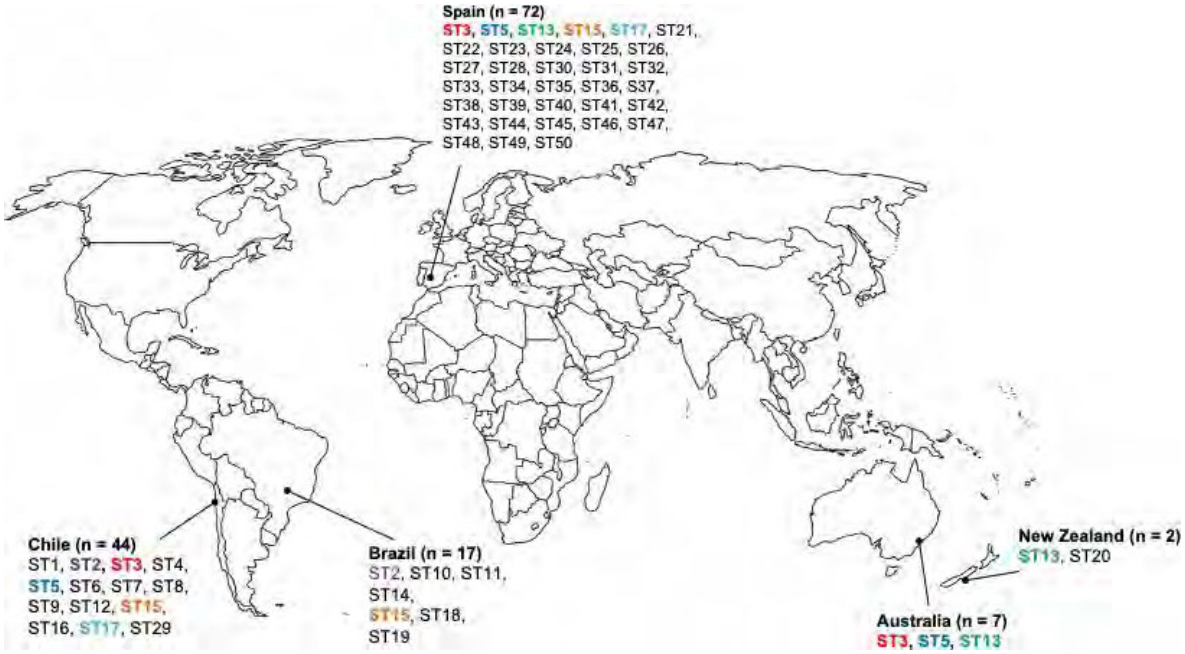


Figure 5.3. Global distribution of the herein identified sequence types obtained from *P. jirovecii* positive clinical samples. Sequence types shared across countries are color coded, and N refers to the total number of sequences identified from the specified country.

5.4.1 The *β-tubulin* locus

β-tubulin is a nuclear locus, coding for an important structural microtubule protein, which was first used in MLST analysis of PCP cases in 2001 and has since been used in a wide array of reports for both genotyping and detection [13, 20, 21]. *β-TUB* has remained relatively conserved with a minimal variability recorded, with only five allele types having been described in the literature, until the recent addition of an entirely new allele type from the Spanish cohort [14]. Previously in European studies, including the surveillance of Portugal, Switzerland, Germany, Belgium, Italy, and France, AT1 and AT2 were overwhelmingly the predominant allele types detected [7, 19, 22-25], while Australian genotyping also confirmed the same [3]. It is important to note that in some publications AT1is referred to as Allele Type A, and AT2 is referred to as Allele Type G, due to the positional change at SNP 282, although this MLST schemes considered two additional SNPs. Previous

reports have been heavily based on samples originated from Europe, and there was no South American genotyping analysis which included the *β-TUB* locus to the best of our knowledge. Considering all the samples submitted to the current ISHAM MLST database for *P. jirovecii* confirmed the previously reported data, with AT1 and AT2 being the most prominent allele types, at 45% and 51%, respectively. AT1 was detected across all five countries, and in the literature is overwhelmingly the most abundant one. There has been no clinical association yet linked with either of the two allele types. Previous Australian genotyping showed only the detection of AT2, while current trends show a shift towards AT1 being the dominant genotype [3]. Excluding AT5, which was a novel allele type described in the Goterris *et al.* study, interestingly, two never-before identified allele types were also detected, with both being specific to Brazil [14]. When looking at the phylogenetic tree of the locus, it is apparentthat there are two main clusters, with the allele types AT4 and AT5 being distinctly separated from the remaining three. It is possible that AT1, AT2 and AT3 share a distant ancestor with AT4 and AT5 and have been endemically circulating within the population just as long, but were not detected earlier, as no surveillance has been undertaken previously in Brazil. Unfortunately, no meta data were available for the Brazilian cohort, so no correlation between the allele types and their potential clinical impact could be assessed. Since therewere three novel allele types detected within this study, it is possible that *β-TUB* is not as conservative as once hypothesized, justifying further population studies to be undertaken to uncover additional variants.

5.4.2 The *Cytochrome b* locus

Cytochrome b is a multicopy gene, appearing up to six times within the genome, and was first used in an MLST scheme in 2002, but has since gained popularity within schemes since it

was published in Maitte *et al.*, 2003, for its high variability and discriminatory power, second only to ITS, which is recognized as the most variable loci [26, 27]. The *CYB* locus may be highly variable, as it is a target of the drug Atovaquone, being used for prophylaxis and treatment of PCP infections, which may result in a higher level of sequence variation [28]. Atovaquone inhibits the cytochrome bc1 complex, which is a respiratory enzyme present in the inner mitochondrial membrane [29]. Atovaquone is a commonly used drug to treat PCP in primarily HIV positive populations, in cases of intolerance to sulfur drugs, for prophylaxis and for treatment of mild cases of PCP. It is also a widely used anti-malarial drug and is used for the treatment of toxoplasmosis [30]. Mutations are noted to occur at the quinol oxidation (Qo) site as Atovaquone ubiquinone is analogous to this site, which is hypothesized to lead todrug resistance [31, 32].

The *CYB* locus appeared to be the most polymorphic on investigated in the current study, describing 11 different allele types. The allele type AT1 was the most prominent one, occurring in almost 48% of all studied samples, and was detected across three continents, which was to be expected, as it has been previously described to the dominant genotype in many European publications from the last several years, occurring in roughly half of the studied samples [19, 22, 33, 34]. Population studies report that AT1 occurs in 40-70% of the *P. jirovecii* positive samples in the community. Our findings reinforced the notion that AT1 is most likely the globally dominant allele type. Despite this, AT1 was not picked up in the isolates from Rio de Janeiro. Since AT1 was detected in many Chilean samples, it is possible that AT1 is in circulation, but not the dominant genotype in South America. Similarity to *β-TUB*, no analysis of *CYB* has been undertaken in the America's, nor in Oceania, so no comparisons to past findings could be made. Due to the lack of surveillance of this genetic locus outside of Europe, much is unknown about the realities of the global spread of its variants in the global circulation.

AT2 which has been commonly reported in European reports occurring from 30 to 60%, was also found in South America, but all together only comprised 7.6% of allele types detected. AT2 has been linked previously with those on PCP prophylaxis.

AT4 which was previously only recorded in minute amounts, discovered in one isolate in China in 2019 [35], and not in any of the European studies, now comprised 15% of all allele types found in the current study. It was also the most widespread allele type, detected inall four countries [35]. When investigating its phylogenetic origin (Figure 5.1), it is likely thatrather than genotype AT4 being a recent *de novo* mutation, it has in fact been in circulationfor much longer than thought, but just had not been described from Europe.

AT6 and AT7 were most recently reported in Turkish genotyping studies and have herein only been seen in Spanish samples, so it is possible that they are only endemic to Europe.

Interestingly, AT8 and AT9 were previously picked up in 2013 in France in minute amounts and subsequently in Turkey and Poland, but also in minute amounts [18, 22, 33], now both were detected in Chile, as well as AT8 being detected in Brazil and Spain in the current study. Similarly, to AT4, it is possible that AT8 and AT9 are not as novel ashypothesized, but simply more endemic within the southern hemisphere. Within the Chilean cohort, AT5 and AT8 were equally just as dominant as AT1, but the cohort also consisted of AT9 and AT10. When observing the dendrogram of the *CYB* locus in Figure 5.1, it is evident that the allele types AT9 and AT10 have developed *de novo* from AT5 and AT8, respectively, which could explain the genotype distribution within Chile. Although Brazil is aneighboring country, the *CYB* allocation is dissimilar.

From the Brazilian cohort, two novel allele types were described, which are AT12 and AT13, all branching off from AT4. AT4 and AT11 were both previously detected in China, and in our study from Brazil, two countries where Atovaquone is used for an array of medical reasons [36]. The AT11 mutation occurs at the Qo site and has also been detected in the La Reunion Island, although they do not have endemic malaria or HIV infections, the patients with samples caring the allele type AT11 did receive Atovaquone for prophylaxis due to hematological malignancy [32]. In Brazil, a country with both endemic HIV and malaria, Atovaquone may be applying pressure on *CYB*, resulting in a higher variability within the gene, which has previously been theorized [28, 37]. The results from analyzing globalsamples have raised lots of questions about the reality of allele distribution and shown the importance of screening outside of standard geographical regions. To confirm theinterpretations from this sample set, further investigations are needed with a wider geographical array of *P. jirovecii* positive samples.

5.4.3 The *mitochondrial 26S* locus

Another mitochondrial gene used within the ISHAM MLST consensus scheme is the *mt26S* locus, also known as *mtLSU* rRNA, which is involved in metabolic mechanisms, and has previously been considered highly informative in *P. jirovecii* typing studies [38, 39]. The locus has 15 copies within the genome and has been used for identification purposes in hospitals and laboratories globally [40], it is also a popular target in genotyping since 1996 due to its high variability [13, 41]. Observation from the current analysis confirms its high variability, with a detection of 11 allele types. The results were similar to previous reports from Europe, the UK, and the US in the last two decades, and AT4 (also referred to as mt85Ain some literature) was the most commonly detected allele type, being detected in four

countries [42, 43]. AT2 (also referred to as mt85T) was the second most reported allele type, occurring in 20%, which is also in line with the majority of publications [44]. Previous reports showed that only the allele type AT2 was detected in Australian *P. jirovecii* samples, which was also detected in two of the samples of the herein studied Australian cohort, but now the allele type AT4 was the dominant genotype circulating within the population [3]. AT3 and AT5 were found in similar levels, (17.5% and 14.4%, respectively). This is an interesting observation, one made also in the Goterris *et al.* study observing only the Spanish cohort, as AT5 was a genotype occurring only infrequently, while AT3 was often the third dominant genotype detected in the communities [45-47]. Depending on the geographical location of previous reports, AT6, AT7 and AT8 are detected in minute amounts [48], a trendremaining within the global context. Conversely, when looking at the population studies of specific countries, such as Turkey, France or Ireland, these three genotypes occur more frequently [22, 33].

Two novel allele types, AT9 and AT10, were detected in the current analysis, both having been identified from Brazilian samples. From the Brazilian cohort, there was no allele types AT4 and AT2 detected, which are the two globally dominant alleles. Based on the dendrograms, AT10 and AT11, are possibly a variant of AT4, and along with the remainder of the reported polymorphisms varying greatly from location to location, implementing, that various environmental factors which are specific to different geographic locations, may be a confiding influence on emerging of novel genotypes. Over time, especially with the use of more specific genotyping techniques, it appears that there is an increasing variability within the *mt26S* genetic locus. The high frequency of mutations detected, as also I steh case for the *CYB* locus, could be connected to their location within the mitochondria, where variants are naturally more occurring than in nuclear genes, such as *β-TUB* and *SOD* [49]. Further global genotyping would be beneficial to better understand the polymorphism of this allele, the

relationship the location and the environment have on frequency of polymorphisms, and if there are any clinical implications of these variations.

5.4.4 The *superoxide dismutase* locus

Superoxide dismutases are metalloenzymes which are ubiquitously present and help protect against reactive oxygen species [50]. Although relatively conservative, with no more than a handful of variants reported, reports have shown correlation of specific allele types with specific clinical associations, and the locus is becoming increasingly popular since its first reported use in MLST studies in 2003 [51]. The two major allele types herein detected were AT1 and AT2 (also referred to as (SOD110C/215T and SOD110T/215T, respectively), which is in line with other population typing reports, from around the world, including reports from the UK, France, Portugal, Belgium, Poland, China, and Zimbabwe [19, 45, 51-54]. Little studies have been undertaken outside of Europe and may be the reason for the lack of variation previously discovered. AT3 was detected only in Spanish samples, which hadpreviously only been detected in France and Portugal in minute amounts [26, 44], showing possible geographical restritions of less dominant allele types. It is important to note that past reports have only reported *SOD* mutations which are non-synonymous, due to the fact that there is no protein change. However, reports have shown that non-synonymous mutations are able to interfere with protein activity by association and influencing splicing processes or translation [55]. Due to this fact, it was important to add all polymorphisms detected, which results in the detection of three novel genotypes, AT4, AT5 and AT6, again all originating from Brazil. When looking at the dendrograms of the *SOD* allele types, it is highly possible that AT4 and AT6 are closely related to and variants of AT1, while AT5 is in very close association with AT2. Further genotyping in South America, especially Brazil, will allow

researchers to ascertain if these allele types are recent and direct mutations, which occur *de novo* due to certain specific environmental and geographical stresses, or if they simply are endemic to the community, but have never been detected before.

It is important to note that the *SOD* allele types have had various nomenclature in the literature, which makes comparisons difficult and complex, and increases the likelihood of error. For example, it is vital to note that AT4 and AT5 in this study are not the same as allele types labelled SOD4 and SOD5 in two other publications, which were not previously unearthed when the allele types from this database were assigned [22, 38]. The addition of isolates to a global database, using a standardised MLST scheme, will allow for easier and more straight forward merging of results from different research groups.

5.4.5 Mixed infections

CYB, *mt26S* and *SOD* have also shown to appear within mixed infections, which is the case when more than one unique allele type is detected in a single sample, which have been detected in almost all genotyping reports. For the current study, only mixed infections detected due to the occurrence in only one SNP, in only one genetic region were included. Sanger sequencing does not allow for the identification of multiple allele types within one sample, so those samples were excluded from the current study. The more specific techniquesare used for sequencing, the higher mixed infection rates have been detected, with rates ranging from 30% to 70% when using traditional Sanger sequencing, and up to 90% when using next generation sequencing [40, 44, 56]. The herein detected rates of mixed alleleswere 16%, which is slightly lower than other MLST reports, but this may be significantly lower than the true reality, as 43 samples were excluded from our analysis based on poor

sequencing quality or more than one mixed SNP being present. It is possible that previously reported low amplification rates were due to mixed infections with several allele types(strains), which made sequencing analysis not possible. Other factors impacting on the detection of mixed alleles may be reliant on fungal loads, especially in the case of low fungal loads where minor mixtures can be missed. It has also been hypothesised that mitochondrial genes have higher rate of mixture [57], which is true within our dataset as the least number ofmixtures were detected in the *SOD* locus, and mixed alleles were not detected in the *β-TUB* locus. So far, due to a variety of biases and factors influencing the detection, no clinical characteristics had been linked with mixed alleles, seeing high mixture percentages in both colonised and PCP infected individuals, as well as in HIV-positive and HIV-negative individuals [57, 58]. There were no clinical or spatiotemporal links in our dataset associated with the mixed infections. It is also unclear if mixed alleles are coinfections at the same time or arise due to high replication rates of the fungi when causing infection [59, 60]. Mixture detections muddy epidemiological investigations and requires greater tweaking of genotypingtechniques to better detect and understand their impact.

5.4.6 High genetic variation amongst the studied population

Sequence type analysis of the isolates also shows high heterogeneity within the global population studied, with a discovery of 49 sequence types identified from 145 isolates. Most sequence types herein identified were endemic to their respective country of origin, and phylogenetic analysis demonstrated many endemic sequence types clustering together, showing close relationships between the groups. ST3, ST13 and ST15 were the most distributed sequence types, each detected across three countries, which can be expected as they are mostly composed of the most distributed allele types. The most detected sequence

type is ST3, comprising 11% of the global isolates, and is exclusively comprised of the four most detected allele types, *β-TUB* AT1, *CYB* AT1, *mt26S* AT4 and *SOD* AT2.

Due to the high allelic variation within the cohorts, it is evident that *P. jirovecii* mutants are influenced by spatiotemporal factors, but it is yet unclear to which extent both time and geography play a role in increasing polymorphisms. Previously, certain allele variations and haplotypes have been linked to clinical characteristics, such as certain genetic variants being linked to resistance, prophylaxis, morbidity, and mortality of PCP, or colonisation and treatment outcomes. At times, the associations have been contradictory in different reports, and unable to replicated in follow-up studies. These associations could not be addressed in this study, due to the lack of available metadata for most the investigated samples.

5.4.7 Study limitations

Unfortunately, outside of the Spanish cohort, very limited clinical or spatiotemporal metadata had been available. This is a major hinderance of this study, as the available metadata was insufficient to draw correlation between variants and clinical parameters. Another downside of this study is the unequal sample sizes between the studied cohorts, ranging from 2 to 71 samples, which could skew the realities of the alleles in circulation. Additionally, smaller sample sizes can also skew the data, and may not be a true representation of the genotypes in circulation. Additional submissions of sequence date from more global samples accompanied with stronger metadata to the ISHAM MLST database will lead to an increased potential of the database to be a leading apparatus in gathering *P. jirovecii* epidemiological data, forming more reliable associations and enabling stronger conclusions.

5.4.8 Conclusions

Overall, this study has set the scene for future epidemiological advances which may be made possible within *P. jirovecii* when applying a universally accepted standardized consensus MLST scheme in association with the herein developed database for global genotypingstudies. *P. jirovecii* genotyping research cannot progress at the same speed as in other fungi,if there is no cohesion and unity between research groups. Utilizing the global database will allow researchers to make their results comparable and rectify all clinical and spatiotemporal associations linked to various allele or sequence types. We implore all research groups who have and will genotype *P. jirovecii* to utilize the ISHAM MLST consensus scheme and the associated database to progress our understanding of the global population genetic structure of *P. jirovecii*.

5.4. References

1. Truong, J. and J.V. Ashurst, *Pneumocystis Jirovecii Pneumonia*, in *StatPearls [Internet]*. 2021, StatPearls Publishing.
2. Szydłowicz, M., *et al.*, *Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients—preliminary report.* Parasitology research, 2019. **118**(1): p.181-189.
3. Phipps, L.M., *et al.*, *Nosocomial Pneumocystis jirovecii pneumonia: lessons from a cluster in kidney transplant recipients.* Transplantation, 2011. **92**(12): p. 1327-1334.
4. Urbancic, K.F., *et al.*, *Taking the challenge: a protocolized approach to optimize Pneumocystis pneumonia prophylaxis in renal transplant recipients.* American journal of transplantation, 2018. **18**(2): p. 462-466.
5. Mori, S. and M. Sugimoto, *Pneumocystis jirovecii pneumonia in rheumatoid arthritis patients: risks and prophylaxis recommendations.* Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine, 2015. **9**: p. CCRPM. S23286.

6. Iriart, X., *et al., Risk factors of Pneumocystis pneumonia in solid organ recipients in the era of the common use of posttransplantation prophylaxis.* American Journal of Transplantation, 2015. **15**(1): p. 190-199.

7. Esteves, F., *et al., Identification of relevant single-nucleotide polymorphisms in Pneumocystis jirovecii: relationship with clinical data.* Clinical microbiology and infection, 2010. **16**(7): p. 878-884.

8. Olsson, M., *et al., Identification of Pneumocystis carinii f. sp. hominis gene sequences in filtered air in hospital environments.* Journal of Clinical Microbiology, 1998. **36**(6): p. 1737-1740.

9. Alanio, A., *et al., Investigating Clinical Issues by Genotyping of Medically Important Fungi: Why and How?* Clinical Microbiology Reviews, 2017. **30**(3): p. 671-707.

10. Matos, O. and F. Esteves, *Pneumocystis jirovecii multilocus gene sequencing: findings and implications.* Future microbiology, 2010. **5**(8): p. 1257-1267.

11. van Hal, S.J., *et al., Clinical significance and phylogenetic relationship of novel Australian Pneumocystis jirovecii genotypes.* Journal of clinical microbiology, 2009. **47**(6): p. 1818-1823.

12. Le Gal, S., *et al., Circulation of Pneumocystis dihydropteroate synthase mutants in France.* Diagnostic Microbiology and Infectious Disease, 2012. **74**(2): p. 119-124.

13. Pasic, L., *et al., Consensus Multilocus Sequence Typing Scheme for Pneumocystis jirovecii.* Journal of Fungi, 2020. **6**(4): p. 259.

14. L. Goterris, L.P., M. Guerrero Murillo, A. Kan, A. Anton, J. Aguilar Company, I. Ruiz-Camps, W. Meyer, M.T. Martin-Gomez, *Pneumocystis jirovecii geneticdiversity in a Spanish tertiary hospital.* . Medical mycology, 2021. **Accepted on the 29/8/2021**.

15. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.

16. Kumar, S., *et al., MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. **35**(6): p. 1547-1549.

17. Meyer, W., *et al., Consensus multi-locus sequence typing scheme for Cryptococcus neoformans and Cryptococcus gattii.* Medical mycology, 2009. **47**(6): p. 561-570.

18. Sokulska, M., *et al., Genotyping of Pneumocystis jirovecii in colonized patients with various pulmonary diseases.* Medical mycology, 2018. **56**(7): p. 809-815.

19. Depypere, M., V. Saegeman, and K. Lagrou, *Typing of Pneumocystis jirovecii by multilocus sequencing: evidence of outbreak?* European Journal of Clinical Microbiology & Infectious Diseases, 2016. **35**(6): p. 911-916.

20. Nahimana, A., *et al., Determination of the Copy Number of the Nuclear rDNA and Beta-tubulin Genes of Pneumocystis carinii f. sp. hominis Using PCRMulticompetitors.* Journal of Eukaryotic Microbiology, 2000. **47**(4): p. 368-372.

21. Edlind, T.D., M.S. Bartlett, and J.W. Smith, *Characterization of the beta-tubulin gene of Pneumocystis carinii.* The Journal of protozoology, 1991. **38**(6): p. 62S-63S.

22. Maitte, C., *et al., Multilocus Sequence Typing of Pneumocystis jirovecii from Clinical Samples: How Many and Which Loci Should Be Used?* Journal of Clinical Microbiology, 2013. **51**(9): p. 2843-2849.

23. Gianella, S., *et al., Molecular evidence of interhuman transmission in an outbreak of Pneumocystis jirovecii pneumonia among renal transplant recipients.* Transplant Infectious Disease, 2010. **12**(1): p. 1-10.

24. Schmoldt, S., *et al., Molecular Evidence of Nosocomial Pneumocystis jirovecii Transmission among 16 Patients after Kidney Transplantation.* Journal of Clinical Microbiology, 2008. **46**(3): p. 966-971.

25. Volpe, G., *et al., Genetic Diversity of Pneumocystis carinii Isolated from Human Immunodeficiency Virus-Positive Patients in Turin, Italy.* Journal of Clinical Microbiology, 2001. **39**(8): p. 2995-2998.

26. Maitte, C., *et al., Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* Journal of clinical microbiology, 2013. **51**(9): p. 2843-2849.

27. Takahashi, T., T. Kanda, and A. Iwamoto, *Genetic diversity of drug targets including dihydropteroate synthase, dihydrofolate reductase and cytochrome b, in Pneumocystis carinii f. sp. hominis isolates in Japan.* Res Commun Mol Pathol Pharmacol, 2002. **112**(1-4): p. 159-76.

28. Kazanjian, P., *et al., Pneumocystis carinii Cytochrome b Mutations Are Associated with Atovaquone Exposure in Patients with AIDS.* The Journal of Infectious Diseases, 2001. **183**(5): p. 819-822.

29. Kessl, J.J., *et al., Molecular Basis for Atovaquone Resistance in Pneumocystis jirovecii Modeled in the Cytochrome bc1Complex of Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2004. **279**(4): p. 2817-2824.

30. Looareesuwan, S., *et al., Malarone (atovaquone and proguanil hydrochloride): a review of its clinical development for treatment of malaria. Malarone Clinical Trials Study Group.* The American journal of tropical medicine and hygiene, 1999. **60**(4): p. 533-541.

31. Singh, Y., *et al., Circulating genotypes of Pneumocystis jirovecii and its clinical correlation in patients from a single tertiary center in India.* European Journal of Clinical Microbiology & Infectious Diseases, 2017. **36**(9): p. 1635-1641.

32. Le Gal, S., *et al., Pneumocystis jirovecii Diversity in Réunion, an Overseas French Island in Indian Ocean.* Frontiers in Microbiology, 2020. **11**(127).

33. Sürgeç, E., *et al., Genotyping of Pneumocystis jirovecii isolates obtained from clinical samples by multilocus sequencing: a molecular epidemiology study conducted in Turkey.* Archives of Microbiology, 2020. **202**(7): p. 1647-1652.

34. Desoubeaux, G., *et al., Epidemiological outbreaks of Pneumocystis jiroveciipneumonia are not limited to kidney transplant recipients: genotyping confirms common source of transmission in a liver transplantation unit.* Journal of clinical microbiology, 2016. **54**(5): p. 1314-1320.

35. Wang, M., *et al., Polymorphisms involving the Pneumocystis jirovecii-related genes in AIDS patients in eastern China.* Infection, Genetics and Evolution, 2019. **75**: p. 103955.

36. Li, X.-L., *et al., Pneumocystis Pneumonia in Liver Transplant Recipients: Clinical Course, Epidemiological Analysis and Risk Factors.* 2020.

37. Argy, N., *et al., Pneumocystis Cytochrome b Mutants Associated With Atovaquone Prophylaxis Failure as the Cause of Pneumocystis Infection Outbreak Among Heart Transplant Recipients.* Clinical Infectious Diseases, 2018. **67**(6): p. 913-919.

38. Esteves, F., *et al., Population structure of Pneumocystis jirovecii isolated from immunodeficiency virus-positive patients.* Infection, Genetics and Evolution, 2010. **10**(2): p. 192-199.

39. Wakefield, A.E., *et al., Amplification of mitochondrial ribosomal RNA sequences from Pneumocystis carinii DNA of rat and human origin.* Molecular and biochemical parasitology, 1990. **43**(1): p. 69-76.

40. Alanio, A., *et al., Diversity of Pneumocystis jirovecii during Infection Revealed by Ultra-Deep Pyrosequencing.* Frontiers in Microbiology, 2016. **7**(733).

41. Keely, S.P. and J.R. Stringer, *Multi-locus genotype switching in Pneumocystis carinii sp. f. hominis: evidence for reinfection.* J Eukaryot Microbiol, 1996. **43**(5): p. 50s.

42. Dimonte, S., *et al., Molecular analysis based on mtLSU-rRNA and DHPS sequences of Pneumocystis jirovecii from immunocompromised and immunocompetent patients in Italy.* Infection, Genetics and Evolution, 2013. **14**: p. 68-72.

43. Vera, C., *et al., High transient colonization by Pneumocystis jirovecii between mothers and newborn.* European Journal of Pediatrics, 2017. **176**(12): p. 1619-1627.

44. Esteves, F., *et al., Pneumocystis jirovecii multilocus genotyping profiles in patients from Portugal and Spain.* Clinical microbiology and infection, 2008. **14**(4): p. 356- 362.

45. Szydłowicz, M., *et al., Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients—preliminary report.* Parasitology Research, 2019. **118**(1): p. 181-189.

46. Singh, Y., *et al., Genetic polymorphisms associated with treatment failure and mortality in pediatric Pneumocystosis.* Scientific Reports, 2019. **9**(1): p. 1192.

47. Sokulska, M., *et al., Genotyping of Pneumocystis jirovecii in colonized patients with various pulmonary diseases.* Medical Mycology, 2017. **56**(7): p. 809-815.

48. Desoubeaux, G., *et al., Epidemiological Outbreaks of Pneumocystis jirovecii Pneumonia Are Not Limited to Kidney Transplant Recipients: Genotyping Confirms Common Source of Transmission in a Liver Transplantation Unit.* Journal of Clinical Microbiology, 2016. **54**(5): p. 1314-1320.

49. Kang, D. and N. Hamasaki, *Mitochondrial oxidative stress and mitochondrial DNA.* 2003.

50. Younus, H., *Therapeutic potentials of superoxide dismutase.* International journal of health sciences, 2018. **12**(3): p. 88-93.

51. Wakefield, A.E., *et al., Limited Asymptomatic Carriage of Pneumocystis jiroveci in Human Immunodeficiency Virus–Infected Patients.* The Journal of Infectious Diseases, 2003. **187**(6): p. 901-908.

52. Esteves, F., *et al., Pneumocystis jirovecii multilocus genotyping in pooled DNA samples: a new approach for clinical and epidemiological studies.* Clinical Microbiology and Infection, 2012. **18**(6): p. E177-E184.

53. Miller, R.F., *et al., Isolates of Pneumocystis jirovecii from Harare show high genotypic similarity to isolates from London at the superoxide dismutase locus.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 2005. **99**(3): p. 202-206.

54. Xue, T., *et al., Genetic Polymorphisms of 3 Cases Pneumocystis jirovecii in Shanxi Province PR China.* 2021.

55. Parmley, J. and L. Hurst, *How do synonymous mutations affect fitness?* BioEssays : news and reviews in molecular, cellular and developmental biology, 2007. **29**: p. 515-9.

56. Nahimana, A., *et al., Typing of Pneumocystis carinii f. sp. hominis by PCR-SSCP to indicate a high frequency of co-infections.* Journal of medical microbiology, 2000. **49**(8): p. 753-758.

57. Alanio, A., *et al., Correlation Between Pneumocystis jirovecii Mitochondrial Genotypes and High and Low Fungal Loads Assessed by Single Nucleotide Primer Extension Assay and Quantitative Real-Time PCR.* Journal of Eukaryotic Microbiology, 2015. **62**(5): p. 650-656.

58. Tsolaki, A.G., *et al., Genetic diversity at the internal transcribed spacer regions of the rRNA operon among isolates of Pneumocystis carinii from AIDS patients with recurrent pneumonia.* Journal of Infectious Diseases, 1996. **174**(1): p. 141-156.

59. Valero, C., *et al., Copy number variation of mitochondrial DNA genes in Pneumocystis jirovecii according to the fungal load in BAL specimens.* Frontiers in microbiology, 2016. **7**: p. 1413.

60. Ma, L. and J.A. Kovacs, *Genetic analysis of multiple loci suggests that mutations in the Pneumocystis carinii f. sp. hominis dihydropteroate synthase gene arose independently in multiple strains.* Antimicrobial agents and chemotherapy, 2001.**45**(11): p. 3213-3215.

Chapter 6

The role of the genetic locus *Dihydropteroate Synthases* (*DHPS*) in *Pneumocystis* research

6.1 Introduction

A common theme among fungi, including *Pneumocystis* spp., is that they are unable to uptake folic acid from their host as mammalian cells can. However, all living cells require folic acid for growth, which is vital for the production of purines, methionine, thymidine and glycine [1]. Therefore, fungi must rely on the *de novo* production, and hence contain folic acid synthesis (FAS) genes, which control the folate biosynthetic pathway [2]. The FAS genes are from a single open reading frame that encodes a trifunctional, multidomain enzyme, which includes the dihydropteroate synthases (*DHPS*) gene, as well as the dihydroneopterin aldolaseand the hydroxymethyldihydropterin pyrophosphokinase genes [3, 4]. DHPS is the initiation enzyme, which is the catalyst for the condensation of para-aminobenzoic acid (PABA)together with 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (DHPPP), forming 7,8- dihydropteroate, which, along with L-glutamate and the enzyme dihydrofolate synthase then changes into 7,8-dihydrofolate. 7,8-dihydrofolate is the precursor for folate, which is then catalysed into tetrahydrofolate by dihydrofolate reductase (DHFR) [2, 4-6]. DHFR is reliant on nicotinamide adenine dinucleotide phosphate in its reduced form to function [7].

The first line of treatment and prophylaxis of *P. jirovecii* relies on the use of trimethoprim-sulfamethoxazole (TMP-SMX), also known as co-trimoxazole. It consists of one part trimethoprim to five parts sulfamethoxazole, and it is used for a wide range of bacterial, fungal and protozoal infections [8]. TMP-SMX can inhibit the folic acid synthesis (FAS), due to the sulfamethoxazoles. Trimethoprim, which works in synergy with sulfamethoxazole, targets dihydrofolate reductase by inhibiting it. Multiple sources of evidence exist that suggest that trimethoprim is ineffective when treating *P. jirovecii*, and that sulfamethoxazoles works in similar fashion as mono-therapy [9]. Sulfonamides are able to do so as they are structurally similar to PABA, and compete with it, to inhibit the DHPS

from catalysing PABA and pteridine, disabling the conversion to dihydropteroic acid. To do this, sulfonamides compete with PABA by inhibiting DHPS from catalysing the condensation of PABA and, which will form dihydropteroic acid. This is visualised in Figure 1.

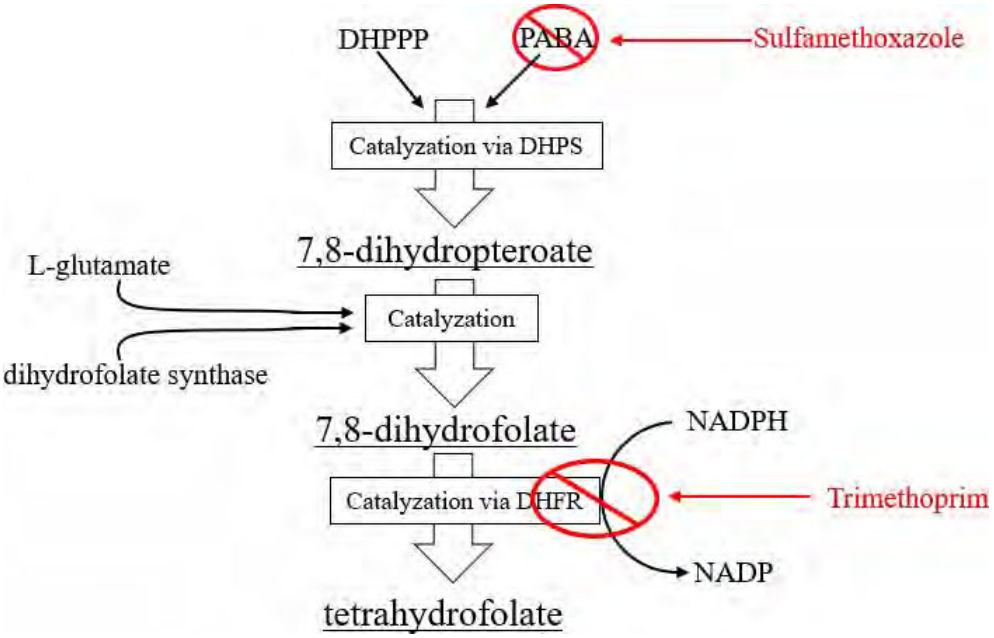


Figure 6.1. The biosynthesis pathway of folic acid synthesis starting with the catalysation of DHPPP and PABA via DHPS, and the mechanism TMP-SMX uses to interfere with the pathway.

When the two enzymes which are targeted by TMP-SMX, DHPS and DHFR, the biosynthesis pathway is inhibited, leading to the breakdown of the fungal cell, since the production of amino acids ceased, which leads to a shortage of protein synthesis [1].

6.1.1 *DHPS* and resistance

The genetic locus *DHPS* became of interest in *P. jirovecii* studies in the 1990's, as HIV was a global epidemic, and PCP was increasingly more common amongst AIDS patients. Towards the end of the decade, TMP-SMX was widespread for the treatment of PCP in AIDS patients,

and genotyping techniques become more increasingly used [10]. The widespread use of sulfonamides caused concern, as even previous short-term use showed the emergence of sulfa resistant bacteria, seen in *Escherichia coli*, *Neisseria meningitidis* and *Plasmodium falciparum* [11, 12]. The number of PCP patients which were unresponsive to TMP-SMZ started to rise during the 90's and it became statistically relevant that a link between sulfa exposure and mutations may be present. Early techniques of genotyping including SSCP and RFLP analysis, as research groups started looking for mutations in isolates from patients who were not responding to PCP treatment [11, 13]. In 1998, the first non-synonymous mutations in the *DHPS* region were detected in codons 55 and 57, which were reflective of a base pair change at SNP 165 and SNP 171 [14]. A wildtype *DHPS* allele has an adenine (A) at the nucleotide 165 and a cytosine (C) at position 171, and variants have a nucleotide change at 165 into a guanine (G), or into thymine (T) at position 171. The SNPs cause amino acid substitutions, leading to an alanine substitution from threonine at codon 55, and a serine substitution from proline at codon 57 [15]. Going forward, any wild type allele is referred to as AT1, a single mutation at codon 55 is AT2, while a single mutation at codon 57 is AT3, and a double mutation at both sites is referred to as AT4. In comparison to the same point mutations within *E. coli*, the SNPs are within an active site involved with substrate binding ofthe enzyme, and strongly indicates that amino acid substitutions could impact the activity, emphasising the importance of the detection of the mutations [16]. This finding was significant, as *DHPS* is generally a highly conserved region, and no mutations had been detected from any *Pneumocystis* species [17]. Kazanjian *et al.,* 1998, detected the mutations and hypothesised that the mutations from HIV positive patients correlated to the use of sulfa prophylaxis up to four months before any PCP diagnosis [14]. These mutations were notnoted prior to the late 90's in any genotyping reports, strongly implying that the use of TMP- SMX as prophylaxis and/or treatment was directly related to the emergence of variants [18].

Since then, *DHPS* has been the focus of over 60 genotyping reports, all outlined in Table 1. In reports following Kazanjian *et al.* many also observed *DHPS* mutations at the same positions and reinforced the link of TMP-SMX driving the increase in variants. Overthe next several years, *DHPS* genotyping detected higher levels of mutations, including a rise in double mutants, and the mutants started to be associated with higher rates of treatment failure in PCP patients, treated with TMP-SMZ and/or dapsone, as well as higher mortality rates [16, 18-20]. Further reports also made associations that patients with mutations correlated significantly with a range of clinical attributes, such as treatment failure, increased hospitalisation, severe disease, longer ventilation, and higher mortality [18, 20-22]. A significant increase of variants with time was closely correlated with the rate of use of sulfa drugs as prophylaxis, but since *P. jirovecii* cannot be cultured *in vitro,* this direct link could not be confirmed. As research groups around the world started screening various *P. jirovecii* positive patients, mutation rates then varied greatly between geographical regions, ranging from under 5% in South Africa to over 80% in the US, and even varying greatly within the same country, as witnessed in France, with detection rates reported from 8 to 40%. Interestingly, in early reports, it were developed countries, such as the US and European countries, which harboured higher levels of mutants, where high sulfa drugs were being used to treat PCP, while lower rates were detected in developing countries or regions, such as under 5% in China, India and Thailand, 4% in South Africa, and 7% in Zimbabwe, where sulfa use was limited [23-25]. From the mid 2000's, the detection of mutations started to fall, while contradictory results and conclusions within the reports started to rise. Mutations were increasingly reported in patients with no prior sulfa exposure, while cohorts who previously had high frequency of mutations, such as HIV positive patients with TMP-SMX prophylaxis/treatment harboured wild type alleles [26, 27]. Subsequent studies also failed to find a links between the mutations and an increased morbidity nor mortality, treatment

failure, increased ventilation or length of hospital stay [28-30]. This led to the belief that mutations may have been acquired *de novo*, alongside being under pressure from exposure to sulfa drugs, which also strengthened the hypothesis that *Pneumocystis* was acquired via airborne transmission, opposed to recurrent infections from childhood [26].

Despite the paradoxes reported and documented in Table 6.1, strong evidence exists that sulfa exposure remains a major influence on mutations with several strong studies supporting this claim. After the original low screening of mutants in South Africa, Dini *et al.* 2010 detected 56% mutations in the population and attributed them directly to the increased HIV prevalence in the country, along with an increase of HIV related TMP-SMX interventions causing selection pressure on the genetic loci [31]. Four years later, the rate in South Africa had increased to 87% [32]. A similar trend was also documented within Uganda, where all isolates surveyed harboured mutants, despite only two given TMP-SMX for PCP treatment. Population analysis showed that sulfa drugs were heavily used for malaria prophylaxis and the heavy selection pressure on the general population resulted in the variants [23]. Additionally, mutations were very rarely ever detected in infants [33], Another study found that patients surveyed in London in 1992 – 1993, and then again in 2001, found that there was a reversal of mutant to wild type rations once the selection pressure from sulfa drugs was removed [34]. It is understood that with the introduction of HAART therapy, alongside the decrease of HIV infections and the limiting of sulfoamides in widespread prophylaxis, mutations were reported less frequently and in the last decade have started to become rarer [35, 36]. The differences of *DHPS* allele types detected over the last three decades and the contradictions reported indicates that further genotyping of this locus is required to expand on our understanding of variants, and their clinical significance. Theherein undertaken analysis aimed to screen the *DHPS* locus from *P. jirovecii* isolates four countries to identify the current situation of the presence of mutations.

Table 6.1. A chronological list of all publications with *P. jirovecii* genotyping analysis of the *DHPS* genetic locus, with a breakdown of mutations located in codons 55 and 57, and research conclusions of the allele from the respective articles. N/A signifies that no data was offered regarding mixed alleles.

Reference	*Country*	*Year of Isolation*	*AT1 wt Thr55/ Pro57*	*AT2 Ala55/ Pro57*	*AT3 Thr55/ Ser57*	*AT4 Ala55/ Ser57*	*Mixed*	*Commentary and Study Conclusions*
Kazanjian 1998 [14]	Ann Arbor, USA	1991–1997	20	1	1	5	1	Observed an increase exposure to Sulfides increases risk of mutations in HIV patients.
Mei, 1998 [19]	Maryland, USA	1998		2*			N/A	Hypothesised patients with mutations may not respond to TMP-SMX. *Paper did not specify which codon the mutations occured.
Helweg-Larsen, 1999 [18]	Copenhagen, Denmark	1989-1989	121	14	8	3	N/A	All HIV-positive cohort. *DHPS* mutation associated with increased death rate, with 3-month survival significantly lower in mutations. Also looked at mutations Codon 60.
Ma, 1999 [16]	Bethesda, USA	1985-1998	18	3	0	9	N/A	Significant association between prior exposure to TMP-SMX or Dapsone and the presence of mutations at codon 55 and/or codon 57. Also looked at Codon 171.
Ma and Kovac, 2000 [37]	Bethesda, USA	1986-1999	5	7	0	7	3	Link between exposure to Sulfa drugs and mutations, but also possible people infected with mutants via airborne transmission
Kazanjian 2000 [20]	Boston, Indianapolis, Detroit, and Denver, USA	1991-1999	55	0	1	40	1	Association between Sulfa prophylaxis exposure and *DHPS* mutations and shows that the incidence of mutations is increased by the duration of Sulfa prophylaxis.
Takahashi, 2000 [22]	Tokyo, Japan	1994-1999	36	4	2	45	0	Patients with mutations did not respond to TMP-SMX.
Huang, 2000 [38]	Atlanta, San Francisco and Seattle, USA	1991-1999	35	6	2	55	13	All HIV positive cohort, found that Sulfa prophylaxis was independent of mutations as found mutations in patients with no prophylaxis history.
Navin, 2001 [39]	Atlanta and San Francisco, USA	1995-1999	36	4	2	45	0	Found patients with mutations can be successfully treated with TMP-SMX.
Visconti, 2001 [40]	Rome, Italy	1992-1997	13	4	0	3	N/A	Mutations are significantly associated with PCP prophylaxis failure, but not significantly associated with sulpha therapy.
Ma, 2002 [41]	Milan, Italy	1994-2001	98	3	2	2	2	Found association between mutations and prior sulfa prophylaxis.

	Atlanta, Los Angeles, San Francisco and Seattle, USA	1996–2000	115	18	0	181	0	Same study used to compare 384 HIV positive adults, and found overwhelming majority were mutations.
Crothers, 2005 [21]	San Fransico, USA	1997-2002	40	0	0	74	N/A	Patients receiving Sulfa prophylaxis within 3 months of PCP had higher rates of mutations. In contrast, patients with HIV diagnosis within 30 days of PCP were less likely to harbour mutations.
Roberts, 2005 [47]	South Africa	2000-2003	51	0	0	1	0	The mutations were not associated with long term prior Sulfa use. Study suggested that geographic area is an independent factor of mutations, due to low local sulfonamide practices.
Riebold, 2006 [48]	Germany	Unknown	71	0	0	0	0	3 patients with the mutation at *DHPS* codon 171.
Wissman, 2006 [49]	Porto Alegre, Brazil	1997-2004	57	0	0	0	N/A	5 cases had sulfa prophylaxis, but no mutations.
Valerio, 2006 [50]	Milan, Italy	1996-2006	154	4	1	4	11	Showed a correlation between mutation and disease severity and death.
Costa, 2006 [51]	Lisbon, Portugal	2001-2004	62	3	0	2	1	The frequency mutations were lower than in previous Portuguese studies.
Esteves, 2008 [36]	Lisbon, Portugal	2001-2004	51	0	0	2	2	A decline of mutation was noted, possibly due to" decreased use of prophylaxis with sulphonamides following the introduction of HAART".
	Seville, Spain	2002-2004	28	2	2	1	3	
Siripattanapipong 2008 [25]	Bangkok, Thailand	1997–2003	15	0	0	2	N/A	Lower rate of *DHPS* mutations that those in Europe or US.
Alvarez-Martínez, 2008 [52]	Spain	2000-2004	181	2	3	1	1	Unable to show that mutations were associated with mortality or poor treatment outcome.
van Hal, 2009 [28]	Sydney, Australia	2001-2007	52	3	5	0	N/A	Data did not find an association between mutants and TMP-SMX prophylaxis but did associate with mutations and severe disease.

Montes-Cano, 2009 [53]	Seville, Spain	N/A	8	0	0	0	0	No mutations noted.
Hauser, 2010 [29]	Lyon, France	1993-1997	130	0	32	9	23	Mutants significantly higher in Lyon, but the study noted there were "no significant differences in age, sex, HIV status, HIV risk factor, and CD4 cell count among patients infected" with mutants or wt.
	Lausanne, Switzerland	1990-2000	93	2	0	2	5	
	Zurich, Switzerland	1994-1998	9	0	0	2	4	
Dini, 2010 [31]	Johannesburg, South Africa		66	4	7	18	31	Found that mutations are likely due to selective pressure in TMP-SMX use, both for PCP and other infections, and associated with reduced sensitivities to Sulfa drugs.
Esteves, 2010 [54]	Lisbon, Portugal	1997–2007	77	3	4	0	3	Study notes the decline of mutations in Portugal.
Jarboui, 2011 [55]	Tunis, Tunisia	2010	15	0	0	0	2	AT2 and AT3 detected in mixture with wt.
Zhu, 2011 [56]	Hangzhou, China	2015-2016	59	1	0	0	0	Entire cohort of AIDS patients with no mutations found.
Phipps, 2011 [57]	Sydney, Australia	2010-2011	20	0	0	0	N/A	No mutations found.
Beser, 2012 [35]	Sweden	1996-2003	199	0	0	0	N/A	Study looks at population in Sweden, looking at 8 years after HAART therapy. Sweden have a restricted use of TMP-SMX when treating diseases other than PCP, that may lead to lack of mutations.
Damiani, 2012 [58]	Amiens, France	Unknown	9	1	2	0	3	Sampled air around patients and found same genotypes in circulation.
Maitte, 2013 [59]	Nantes, France	2006-2011	33	0	0	0	N/A	No mutations found.
Rabodonirina, 2013 [60]	Lyon, France	1993-1996	75	0	14	20	N/A	Found that mutations had no effect on mortality rate within 1 month.
Dimonte, 2013 [61]	Rome, Italy	2006 - 2010	21	0	0	0	N/A	No mutations found - support the hypothesis that also mutants can be transmitted through infected hospital visitors.
Yoon, 2013 [62]	San Francisco, USA	1997-2008	69	6	4	159	63	Demonstrate that prophylaxis, including dapsone and not just TMP-SMX, increased the risk of mutant,

Long, 2014 [63]	Beijing, China	2008-2011	22	0	0	0	0	HIV negative PCP patients, no prophylaxis given, and no mutations noted.
Deng, 2014 [64]	Guangzhou, China	2009-2013	22	0	2	1	N/A	Confirmation of low mutation rates through China, no prophylaxis in patients.
Mogoye, 2014 [32]	South Africa	2008-2010	85	16	30	165	329	Mutant genotype in total make up 87%. High mutation prevalence found but no statistically significant associations and prior sulfa drug exposure, mortality, or HIV status.
Kim, 2015 [65]	Seoul, South Korea	2007-2013	129	0	0	0	N/A	No mutations despite prophylaxis use.
Le Gal, 2015 [66]	Brest, France	2002-2012	12	0	0	0	1	Only isolate with mutation had history of TNX-SMX prophylaxis.
Frizza, 2015 [67]	Seville, Spain	Unknown	54	5	5	3	8	This study looked at two cohorts, PCP and colonised. None of the colonized patients received sulfa drugs in the 6 months before isolation and no mutations noted, while 52% of PcP patients had received TMP-SMX, and 6 had mutations.
Suárez, 2017 [68]	Cologne, Germany	2011-2016	39	0	1	1	0	Mutations also noted in codon 171.
Ponce, 2017 [30]	Santiago, Chile	2005-2007	29	3	1	8	14	Observed no association between mutants and mortality. Study found high frequency of mutations (48%) in first time PCP patients, with no sulfa history.
Nevez, 2017 [69]	Brittany, France	2005	86	0	0	0	N/A	All cystic fibrosis cohort, no mutations despite high use of prophylaxis.
Montesinos, 2017 [70]	Brussels, Belgium	Unknown	68	0	1	1	1	No association with prior sulfa use.
Ricci, 2018 [71]	São Paulo, Brazil	2011	30	0	0	0	N/A	No mutations found.
Ozkoc, 2018 [72]	Izmir, Turkey	2011-2014	28	0	0	0	N/A	No mutations found.
Szydłowicz, 2019 [73]	Wroclaw, Poland	2015-2018	5	0	0	0	N/A	No mutations found but all receiving prophylaxis in the past.
Singh, 2019 [74]	New Delhi, India	Unknown	30	0	0	0	N/A	All paediatric patients. 21 of 30 had mutations at codons 96 and 98.
Muñoz, 2020 [75]	Medellin, Colombia	2004-2005	42	2	1	0	N/A	3 patients with mutations had TMP-SMX prophylaxis in 4 months prior to PCP detection.

Taylor, 2021 [76]	Kampala, Uganda	2007-2008	0	0	10	3	2	High mutation rate, but all patients with mutations denied prior prophylaxis to PCP, but high levels of TMP-SMX
Gurbuz, 2021 [77]	Izmir, Turkey	2016-2019	135	0	0	0	2	Both mutations are mixed infections with wt and AT3 in two hospitalised cases but believe sulfa pressure did not cause mutation.
Tsvetkova, 2021 [78]	Bulgaria	Unknown	30	0	0	0	N/A	No mutations found.
Nawel, 2021 [79]	Tunis, Tunisia	2016 & 2018	2	0	0	0	N/A	Looked at two renal transplant patients.
Goterris, 2021	Barcelona, Spain	2014-2018	71	0	0	0	0	No mutations detected.

6.2 Methods

6.2.1 *Pneumocystis jirovecii* positive Samples and DNA Extraction

A total of 84 samples were obtained from four countries, including Australia, Brazil, Chile, and New Zealand, isolated between 2016 and 2018. A detailed list of the samples is shown in Appendix Table 1. Samples were previously confirmed to be positive for *P. jirovecii* and were sent to the Medical Mycology Research Laboratory for genotyping. The specimens were not from active infections of patients. DNA was extracted from Australian isolates using the QIAamp DNA Mini Kit as instructed, with the inclusion of a liquid nitrogen grinding step, as demonstrated in Chapter 2.

6.2.2 PCR Amplification

For the identification of *P. jirovecii* the genetic locus *DHPS* was amplified by PCR using the forward 5´-AGCAGTGCCCCAAATCCTAT-3´ primer, and reverse primer 5´-GCGCCTACACATATTATGGCCATTTTAAATC-3´. Samples were amplified in volumes of 25 µl per PCR reaction, using 10 X buffer (100 mM Tris-HCL, pH 8.3, 500 mM KCl, 15 mM MgCl2 and 0.01% w/v gelatine), 50 nM MgCl2, 2 mM dNTPs, 10 ng/µl of each primer, 5 U/µl BIOTAQ DNA (Bioline) polymerase and 10 µl of genomic DNA. Amplification conditions were 95°C for 3 minutes for the initial denaturing step; followed by 45 cycles of 94°C for 30 seconds, 55°C for 45 seconds and 72°C for 45 seconds; and finalised with 72°C for 7 minutes. Negative controls were used got detection of any possible contaminations present.

6.2.3 PCR Visualization

All PCR amplicons were visualized via 1.5% agarose gel electrophoresis, prepared in 100 ml of TBE buffer (1 M Tris-HCL, 1 M Boric acid, 0.02 M EDTA). 1.5 µl of Ethidium bromide/100 µl agarose solution was added and mixed well. 1 µl of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in water) was added to 5 µl ofeach PCR sample, and 1 Kb Plus DNA Ladder (Invitrogen) was added for reference. The obtained fragments were separated via electrophoresis at 80 V for 40 minutes, and then visualized under UV illumination on the transilluminator.

6.2.4 DNA Sequencing and Genotype Analysis

Bidirectional sequencing was performed at Macrogen Inc., Seoul, South Korea. The obtained sequences where then assembled and edited using the software package Sequencher ver. 5.4.6 (Gene Codes Corporation). The cleaned-up sequences for each locus were aligned with the program ClustalW [80] part of the software MEGA version 10.1 [81] and compared to reference sequences found on the ISHAM database at http://mlst.mycologylab.org/pjirovecii/. Allele types were named with respect to the previously published nomenclature on thedatabase.

6.3 Results

The *DHPS* locus was successful amplified and sequenced from all samples. From the 84 *P. jirovecii* positive samples studied, 77 had the *DHPS* allele type AT1 (wild type), 3 had AT3, and 3 had AT4 (Table 6.2). No sample with the allele type AT2 was detected. There were no

mixed alleles detected within the samples, even in isolates which were shown to have mixed alleles for other genetic loci. Additionally, two synonymous point mutations were detected but there was no amino acid change noted, leading to variation 1 of the wild type (AT1 v1), which was identified in a Chilean *P. jirovecii* isolate, and variation 2 of the wild type (AT1 v2), which was identified in a Brazil *P. jirovecii* isolate. In one isolate from New Zealand, one nucleotide insertion between SNP 202 and 203 was also noted (AT1 v3), but there was no amino acid change displayed. All variants detected, and their respective SNPs polymorphisms are displayed in Figure 6.2.

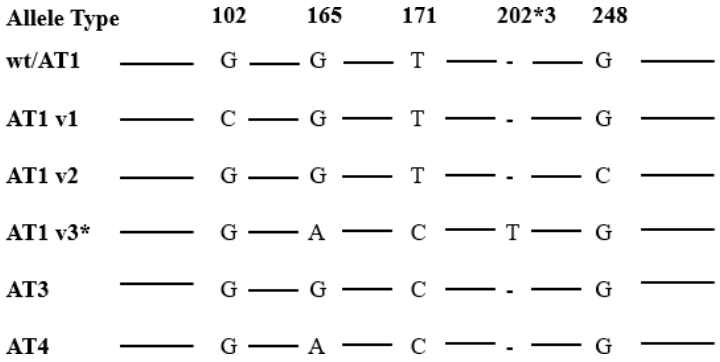


Figure 6.2. The SNP of polymorphisms within the DHPS locus detected in the cohort investigated. AT1, AT3 and AT4 are internationally recognised, and deemed important for insight into TMP-SMX resistance. Wild type variants (AT1 v1-3) were also discovered, but do not harbour any non-synonymous mutations. NB:* AT1 v3 has an insertion between positions 202 and 203. The bp numbers given correspond to the AT1/wildtype.

Table 6.2. The distribution of *DHPS* alleles within the cohort examined.

Allele Type	**Country**	**Rate**
AT1	Australia	12/12 (100%)
	Brazil	19/20 (95%)
	Chile	45/50 (90%)
	New Zealand	2/2 (100%)
AT3	Chile	2/50 (4%)
	Brazil	1/20 (5%)
AT4	Chile	3/50 (6%)

6.4. Discussion

In 1998 mutations were discovered in HIV patients with failing PCP treatment, and it was shown that the mutations influenced the sulfa binding site and reduced the sensitivity to sulfa treatments. Since then, the analysis of codons 55 and 57 has been prevalent in *Pneumocystis* epidemiology, and despite the seeming paradox behind the mutations, there is still no precise explanation for the observations. It is also unclear which geographical factors play an important role on the various mutation rates, but the consensus within the literature is that intrinsic epidemiological factors unique to certain countries/locations, along with interhuman transmission and the sulfa pressure, all play a role in the allele patterns identified, but their extent remains undecided.

The herein conducted study analysed the *DHPS* region of isolates from four countries, between 2016 and 2018. These samples are relatively recent, compared to the data reported in most of the literature dating from the 2000 – 2010, providing information for the distribution of variants two decades after their detection, and the results are outlined in Table 6.2.

No non-synonymous mutations were evident within the Australian cohort nor from the two isolates from New Zealand. This is in line with the last report from Australia, from 2011, which also failed to detect any variants, although mutations were detected in low levels in the early 2000's [28, 57]. All isolates came from non-HIV patients, and were from within the renal ward, and had not received any sulfa-based prophylaxis within the last 6 months. Unfortunately, there are no previous reports from New Zealand, and no metadata available for further insights or comparisons. Both countries have some of the world's lowest rates of HIV infections within developed countries, limiting the use of widespread PCP prophylaxis. As such, the lack of widespread TMP-SMX use is most likely not exerting enough selective pressure within the circulating genotypes to promote variants.

The remaining two cohorts are both from South America but have very differenthistories of mutations detected. Two genotyping studies have been undertaken previously,and there were no variants detected previously, even from patients with recent sulfa prophylaxis [49, 71]. This study found a mutation in one of only twenty isolates collected from Rio de Janeiro, Brazil in 2017, with unfortunately no metadata available. Brazil saw an increase of over 20% in HIV infections between 2010 and 2018, comprising 40% of all HIV infection in South America, with rates varying widely from region to region [82]. It ispossible that Brazil has previously had circulating mutations but outside of the regionsscreened and were subsequently missed, or the HIV surge in the past decade has started applying selection pressure which has not been present before. Conversely, in the neighbouring country Chile, the rate of mutations have seemingly dropped since the reported isolates from 2005 - 2007 [83]. Ponce *et al.*, 2017, reported a high frequency of mutations, found in 48% of the isolates, interestingly from patients who did not receive sulfa prophylaxis. Our study investigated fifty samples from HIV positive adults, with mutations being only noted in 10% of the isolates analysed. It is important to note that both countries only harboured single mutations, and no double mutations were detected, which comprised most variants which had been noted a decade prior. Chile has similar sulfa drug usage as neighbouring South American countries, but variant rates remain higher, strongly suggesting there are other factors currently undetermined influencing these polymorphisms. As themetadata regarding the sulfa use was unfortunately lacking, no conclusions could be made regarding clinical references, as this was purely an observational study.

DHPS and *DHFR* are the only two genetic loci where only non-synonymous mutations are recorded when discussing polymorphisms in *P. jirovecii* genotyping reports. Synonymous mutations are almost never included in the genetic analysis of the *DHPS* locus, which is different to the genotyping of other loci traditionally examined for *P. jirovecii.* If a *DHPS*

locus is determined not to have any amino acid changes at codon 55 and 57, then it is classified a wild type allele. Due to this reason, even if autapomorphic alleles are identified, *DHPS* is not adequate phylogenetically informative in MLST schemes, harbouring no discriminatory power with a H-index of 0 reported in the literature [59].

Polymorphic mutations not including codons 55 and 57 were though reported in the US, from Ma *et al.,* 1999, at codon 171, and then again noted in 2006 and 2017 in Germany [48, 68]. The reported point mutation in codon 171, has no amino acid modification, and did not correlate to any clinical significance. More recently, studies from India reported in Singh *et al.*, 2019, did not find any mutations at codon 55 and 57, but 70% harboured mutations at codons 96 and 99 [74]. The G to A point mutation at SNP 288 results in an amino acid change from valine to isoleucine at codon 96, and the G to C change at SNP 294 changes glutamic acid to glutamine at codon 98. Patients who harboured these mutations were significantly associated with severe disease, including the need for ICU hospitalisation and mechanical ventilation. It was also noted that those with double mutants failed PCP treatment and 53% resulted in death. Although this study found no other non-synonymous mutations, synonymous or silent mutations were noted in three isolates. If *DHPS* continues to be a highly studied and sequenced genetic locus of *Pneumocystis*, it will be beneficial to expand the range of reported polymorphisms, to include all synonymous and non-synonymous mutations noted outside of codon 55 and 57 and allocated an allele type. This would bothgive insight into any other variants which may be susceptible to antibiotic selection pressure and their clinical relevance, allowing clinicians to detect any emerging resistance earlier. Additionally, including a wider array of allele types would improve the discriminatory power of the locus, serving also to be phylogenetically informative.

Another interesting aspect of this study was that all samples were successfully amplified and subsequently successfully sequenced. Amplification failure has been reportedin multiple settings, but for the majority, the PCR protocol, and primers they used was complex and time consuming, involving two separate amplification cycles within the one protocol. The traditionally used PCR protocol had herein been optimised, with new primers and conditions published, leading to a significant rise in sequencing success, see Chapter 3. Traditionally multi-copy genes, such as the *mt26S* locus, have been widely used for identification in the past, but more recently *DHPS* has been considered as potential alternative locus for species identification [70]. Future research groups could consider using *DHPS* for both identification and screening of mutation changes, when MLST is not possible,which could be subsequently deposited in a global database for comparative studies. ISHAM has a dedicated database for six genetic loci used for *P. jirovecii* genotyping, including*DHPS*, along with *β-TUB*, *CYB*, ITS1/2, *mt26S* and *SOD*. When and where possible, routine screening and typing of confirmed *P. jirovecii* infections and carriers, and subsequent depositing of the obtained *DHPS* sequences, would allow for real time surveillance of any potential resistance, with all variants could be detected, readily reported, and further investigated on a globally comparative level. The pitfalls of many studies, including the current one, is that the cohort sizes are not extensive and may not adequately represent the genotypes within the community. As review articles of *DHPS* mutations and rates of mutations are rarely undertaken, wider and more frequent surveillance via a global database would allow to explore the impact of *DHPS* mutations in the clinical setting with greater reliability and significance than the smaller cohort analyses allow.

As *DHPS* mutations, their rates and their clinical impact remains to be a paradox when comparing all the available literature, this only further highlights the importance of continuing to survey the *DHPS* locus from *P. jirovecii* positive individuals. There is still no

obvious explanation for the observations and patterns noted within the last three decades, so variants cannot be solely attributed to TMP-SMX exposure. Additionally, there remain many potential influencing cofactors that are yet to be jointly investigated, as no one study has been able to cover such a wide range. Standardizing and pooling together results from smaller cohorts will be necessary for these issues to be addressed comprehensively.

Although there does not currently seem to be a rise in AT2-4 allele types in the populations studied, non-synonymous mutations now detected outside of codon 55 and 57 with clinical significance have been noted. Future studies should consider depositing sequences of positive cases with metadata available in a comparative manner, such as utilising a global database, to eventually have sufficient data to identify predictors andclinical implications of all allele types to be accurately assessed, to detect and monitoremerging resistance.

6.5 References

1. Armstrong, W., S. Meshnick, and P. Kazanjian, *Pneumocystis carinii mutations associated with sulfa and sulfone prophylaxis failures in immunocompromised patients.* Microbes and infection, 2000. **2**(1): p. 61-67.

2. Brooks, D.R., *et al.*, *Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine.* European Journal of Biochemistry, 1994. **224**(2): p. 397-405.

3. Iliades, P., S.R. Meshnick, and I.G. Macreadie, *Mutations in the Pneumocystis jirovecii DHPS gene confer cross-resistance to sulfa drugs.* Antimicrobial agents and chemotherapy, 2005. **49**(2): p. 741.

4. Volpe, F., *et al.*, *The multifunctional folic acid sythesis fas gene of Pneumocystis carinii appears to encode dihydropteroate synthase and hydroxymethyldihydropterin pyrophosphokinase.* Gene, 1992. **112**(2): p. 213-218.

5. Volpe, F., S.P. Ballantine, and C.J. Delves, *Two domains with amino-acid sequence similarity are required for dihydroneopterin aldolase function in the multifunctional folic acid synthesis fas protein of Pneumocystis carinii.* Gene, 1995. **160**(1): p. 41-46.

6. Iliades, P., *et al.*, *Cloning of the Pneumocystis jirovecii trifunctional FAS gene and complementation of its DHPS activity in Escherichia coli.* Fungal Genetics and Biology, 2004. **41**(12): p. 1053-1062.

7. Cody, V., *et al.*, *New insights into DHFR interactions: Analysis of Pneumocystiscarinii and mouse DHFR complexes with NADPH and two highly potent 5-(ω-carboxy(alkyloxy) trimethoprim derivatives reveals conformational correlations with activity and novel parallel ring stacking interactions.* Proteins, structure, function, and bioinformatics, 2006. **65**(4): p. 959-969.

8. Hughes, W.T., S. Feldman, and S. Sanyal, *Treatment of Pneumocystis carinii pneumonitis with trimethoprim-sulfamethoxazole.* Canadian Medical Association Journal, 1975. **112**(13 Spec No): p. 47.

9. Huang, L., *et al.*, *Dihydropteroate synthase gene mutations in Pneumocystis and sulfa resistance.* Emerging infectious diseases, 2004. **10**(10): p. 1721.

10. Martin, J.N., *et al.*, *Emergence of trimethoprim-sulfamethoxazole resistance in the AIDS era.* Journal of Infectious Diseases, 1999. **180**(6): p. 1809-1818.

11. Triglia, T., *et al.*, *Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in Plasmodium falciparum.* Proceedings of the National Academy of Sciences, 1997. **94**(25): p. 13944-13949.

12. Fermer, C., *et al.*, *Sulfonamide resistance in Neisseria meningitidis as defined by site-directed mutagenesis could have its origin in other species.* Journal of Bacteriology, 1995. **177**(16): p. 4669-4675.

13. Goodwin, S.D., *Pneumocystis carinii pneumonia in human immunodeficiency virus-infected infants and children.* Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 1993. **13**(6): p. 640-646.

14. Kazanjian, O., *et al.*, *Dihydropteroate synthase mutations in Pneumocystis carinii associated with sulfa prophylaxis failures.* AIDS, 1998. **12**: p. 873-78.

15. Lane, B.R., *et al.*, *Dihydropteroate synthase polymorphisms in Pneumocystis carinii.* The Journal of infectious diseases, 1997: p. 482-485.

16. Ma, L., *et al.*, *Pneumocystis carinii dihydropteroate synthase but not dihydrofolate reductase gene mutations correlate with prior trimethoprim-sulfamethoxazole or dapsone use.* Journal of Infectious Diseases, 1999. **180**(6): p. 1969-1978.

17. Huang, L., *et al.*, *HIV-Associated Pneumocystis Pneumonia.* Proceedings of the American Thoracic Society, 2011. **8**(3): p. 294-300.

18. Helweg-Larsen, J., *et al.*, *Effects of mutations in Pneumocystis carinii dihydropteroate synthase gene on outcome of AIDS-associated P carinii pneumonia.* The Lancet, 1999. **354**(9187): p. 1347-1351.

19. Mei, Q. and S. Gurunathan, *Failure of co-trimoxazole in Pneumocystis carinii infection and mutations in dihydropteroate synthase gene.* The Lancet, 1998. **351**(9116): p. 1631-1632.

20. Kazanjian, P., *et al.*, *Pneumocystis carinii mutations are associated with duration of sulfa or sulfone prophylaxis exposure in AIDS patients.* The Journal of infectious diseases, 2000. **182**(2): p. 551-557.

21. Crothers, K., *et al.*, *Severity and outcome of HIV-associated Pneumocystis pneumonia containing Pneumocystis jirovecii dihydropteroate synthase genemutations.* Aids, 2005. **19**(8): p. 801-805.

22. Takahashi, T., T. Kanda, and A. Iwamoto, *Genetic diversity of drug targets including dihydropteroate synthase, dihydrofolate reductase and cytochrome b, in Pneumocystis carinii f. sp. hominis isolates in Japan.* Res Commun Mol Pathol Pharmacol, 2002. **112**(1-4): p. 159-76.

23. Miller, R.F., *et al.*, *Genotypes of Pneumocystis jiroveci Isolates Obtained in Harare, Zimbabwe, and London, United Kingdom.* Antimicrobial Agents and Chemotherapy, 2003. **47**(12): p. 3979-3981.

24. Robberts, F.J.L., *et al.*, *Dihydropteroate Synthase and Novel Dihydrofolate Reductase Gene Mutations in Strains of Pneumocystis jirovecii from South Africa.* Journal of Clinical Microbiology, 2005. **43**(3): p. 1443-1444.

25. Siripattanapipong, S., *et al.*, *Study of DHPS and DHFR genes of Pneumocystis jirovecii in Thai HIV-infected patients.* Medical Mycology, 2008. **46**(4): p. 389-392.

26. Latouche, S., *et al.*, *Pneumocystisjirovecii dihydropteroate synthase genotypes in French patients with pneumocystosis: a 1998–2001 prospective study.* Medical Mycology, 2003. **41**(6): p. 533-537.

27. Calderón, E., *et al.*, *Pneumocystis jiroveci isolates with dihydropteroate synthase mutations in patients with chronic bronchitis.* European Journal of Clinical Microbiology and Infectious Diseases, 2004. **23**(7): p. 545-549.

28. van Hal, S.J., *et al.*, *Clinical significance and phylogenetic relationship of novel Australian Pneumocystis jirovecii genotypes.* Journal of clinical microbiology, 2009. **47**(6): p. 1818-1823.

29. Hauser, P.M., *et al.*, *Interhuman Transmission as a Potential Key Parameter for Geographical Variation in the Prevalence of Pneumocystis jirovecii Dihydropteroate Synthase Mutations.* Clinical Infectious Diseases, 2010. **51**(4): p. e28-e33.

30. Ponce, C.A., *et al.*, *High Prevalence of Pneumocystis jirovecii Dihydropteroate Synthase Gene Mutations in Patients with a First Episode of Pneumocystis*

Pneumonia in Santiago, Chile, and Clinical Response to Trimethoprim-Sulfamethoxazole Therapy. Antimicrobial Agents and Chemotherapy, 2017. **61**(2): p. e01290-16.

31. Dini, L., *et al.*, *High Prevalence of Dihydropteroate Synthase Mutations in Pneumocystis jirovecii Isolated from Patients with Pneumocystis Pneumonia in South Africa.* Journal of Clinical Microbiology, 2010. **48**(6): p. 2016-2021.

32. Mogoye, B.K., *et al.*, *Characterisation of Pneumocystis jirovecii DHPS genotypes using a simple, inexpensive restriction fragment length polymorphism analysis.* Southern African Journal of Infectious Diseases, 2015. **30**(2): p. 46-50.

33. Miller, R.F., *et al.*, *Isolates of Pneumocystis jirovecii from Harare show high genotypic similarity to isolates from London at the superoxide dismutase locus.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 2005. **99**(3):p. 202-206.

34. Beard, C.B., *et al.*, *Genetic Differences in Pneumocystis Isolates Recovered from Immunocompetent Infants and from Adults with AIDS: Epidemiological Implications.* The Journal of Infectious Diseases, 2005. **192**(10): p. 1815-1818.

35. Beser, J., *et al.*, *Absence of dihydropteroate synthase gene mutations in Pneumocystis jirovecii isolated from Swedish patients.* Medical Mycology, 2012. **50**(3): p. 320-323.

36. Esteves, F., *et al.*, *Pneumocystis jirovecii multilocus genotyping profiles in patients from Portugal and Spain.* Clinical microbiology and infection, 2008. **14**(4): p. 356-362.

37. Ma, L. and J.A. Kovacs, *Genetic analysis of multiple loci suggests that mutations in the Pneumocystis carinii f. sp. hominis dihydropteroate synthase gene arose independently in multiple strains.* Antimicrobial agents and chemotherapy, 2001. **45**(11): p. 3213-3215.

38. Huang, L., *et al.*, *Sulfa or sulfone prophylaxis and geographic region predict mutations in the Pneumocystis carinii dihydropteroate synthase gene.* The Journal of infectious diseases, 2000. **182**(4): p. 1192-1198.

39. Navin, T.R., *et al.*, *Effect of mutations in Pneumocystis carinii dihydropteroate synthase gene on outcome of P carinii pneumonia in patients with HIV-1: a prospective study.* The Lancet, 2001. **358**(9281): p. 545-549.

40. Visconti, E., *et al.*, *Mutations in dihydropteroate synthase gene of Pneumocystis carinii in HIV patients with Pneumocystis carinii pneumonia.* Int J Antimicrob Agents, 2001. **18**(6): p. 547-51.

41. Ma, L., *et al.*, *Mutations in the dihydropteroate synthase gene of human-derived Pneumocystis carinii isolates from Italy are infrequent but correlate with prior sulfa prophylaxis.* Journal of Infectious Diseases, 2002. **185**(10): p. 1530-1532.

42. Nahimana, A., *et al.*, *Association between a specific Pneumocystis jiroveci dihydropteroate synthase mutation and failure of pyrimethamine/sulfadoxine prophylaxis in human immunodeficiency virus–positive and–negative patients.* The Journal of infectious diseases, 2003. **188**(7): p. 1017-1023.

43. Zingale, A., *et al.*, *Detection of Pneumocystis carinii and Characterization of Mutations Associated with Sulfa Resistance in Bronchoalveolar Lavage Samples from Human Immunodeficiency Virus-Infected Subjects.* Journal of Clinical Microbiology, 2003. **41**(6): p. 2709-2712.

44. Costa, M.C., *et al.*, *Mutations in the dihydropteroate synthase gene of Pneumocystis jiroveci isolates from Portuguese patients with Pneumocystis pneumonia.* International Journal of Antimicrobial Agents, 2003. **22**(5): p. 516-520.

45. Zar, H., *et al.*, *Prevalence of dihydropteroate synthase mutants in HIV-infectedSouth African children with Pneumocystis jiroveci pneumonia.* Clinical infectious diseases, 2004. **39**(7): p. 1047-1051.

46. Totet, A., *et al.*, *Pneumocystis jiroveci dihydropteroate synthase genotypes in immunocompetent infants and immunosuppressed adults, amiens, France.* Emerging infectious diseases, 2004. **10**(4): p. 667.

47. Robberts, F., *et al.*, *Dihydropteroate synthase and novel dihydrofolate reductase gene mutations in strains of Pneumocystis jirovecii from South Africa.* Journal of clinical microbiology, 2005. **43**(3): p. 1443-1444.

48. Riebold, D., *et al.*, *Pneumocystis jiroveci dihydropteroate synthase gene mutations at codon 171 but not at codons 55 or 57 detected in Germany.* Clinical infectious diseases, 2006. **42**(4): p. 582-583.

49. Wissmann, G., *et al.*, *Absence of dihydropteroate synthase mutations in Pneumocystis jirovecii from Brazilian AIDS patients.* Journal of Eukaryotic Microbiology, 2006. **53**(4): p. 305-307.

50. Valerio, A., *et al.*, *DHPS-Mutated Isolates of Pneumocystis jirovecii from HIV-Infected Individuals: Analysis of Related ITS Genotypes.* Journal of Eukaryotic Microbiology, 2006. **53**: p. S108-S109.

51. Costa, M.C., *et al.*, *Genetic characterization of the dihydrofolate reductase gene of Pneumocystis jirovecii isolates from Portugal.* Journal of Antimicrobial Chemotherapy, 2006. **58**(6): p. 1246-1249.

52. Alvarez-Martínez, M.J., *et al.*, *Pneumocystis jirovecii pneumonia in Spanish HIV-infected patients in the combined antiretroviral therapy era: prevalence of dihydropteroate synthase mutations and prognostic factors of mortality.* Diagnostic Microbiology and Infectious Disease, 2008. **62**(1): p. 34-43.

53. Montes-Cano, M.A., *et al.*, *Vertical transmission of Pneumocystis jirovecii in humans.* Emerging infectious diseases, 2009. **15**(1): p. 125.

54. Esteves, F., *et al.*, *Identification of relevant single-nucleotide polymorphisms in Pneumocystis jirovecii: relationship with clinical data.* Clin Microbiol Infect, 2010. **16**(7): p. 878-84.

55. Jarboui, M.A., *et al.*, *Dihydropteroate synthase gene mutations in Pneumocystis jiroveci strains isolated from immunocompromised patients.* Pathol Biol (Paris), 2011. **59**(4): p. 222-5.

56. Zhu, M., N. Ye, and J. Xu, *Clinical characteristics and prevalence of dihydropteroate synthase gene mutations in Pneumocystis jirovecii-infected AIDS patients from low endemic areas of China.* Plos one, 2020. **15**(9): p. e0238184.

57. Phipps, L.M., *et al.*, *Nosocomial Pneumocystis jirovecii pneumonia: lessons from a cluster in kidney transplant recipients.* Transplantation, 2011. **92**(12): p. 1327-1334.

58. Damiani, C., *et al.*, *Possible nosocomial transmission of Pneumocystis jirovecii.* Emerging infectious diseases, 2012. **18**(5): p. 877-878.

59. Maitte, C., *et al.*, *Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* Journal of clinical microbiology, 2013. **51**(9): p. 2843-2849.

60. Rabodonirina, M., *et al.*, *Pneumocystis jirovecii genotype associated with increased death rate of HIV-infected patients with pneumonia.* Emerging infectious diseases, 2013. **19**(1): p. 21-186.

61. Dimonte, S., *et al.*, *Molecular analysis based on mtLSU-rRNA and DHPS sequences of Pneumocystis jirovecii from immunocompromised and immunocompetent patients in Italy.* Infect Genet Evol, 2013. **14**: p. 68-72.

62. Yoon, C., *et al.*, *Dihydropteroate synthase mutations in Pneumocystis pneumonia: impact of applying different definitions of prophylaxis, mortality endpoints and mutant in a single cohort.* Medical Mycology, 2013. **51**(6): p. 568-575.

63. Long, Y., *et al.*, *Pneumocystis jirovecii dihydropteroate synthase gene mutations in a group of HIV-negative immunocompromised patients with Pneumocystis pneumonia.* Experimental and therapeutic medicine, 2014. **8**(6): p. 1825-1830.

64. Deng, X., *et al.*, *Mutational Analysis of Pneumocystis jirovecii Dihydropteroate Synthase and Dihydrofolate Reductase Genes in HIV-Infected Patients in China.* Journal of Clinical Microbiology, 2014. **52**(11): p. 4017-4019.

65. Kim, T., *et al.*, *Clinical characteristics of hospital-onset Pneumocystis pneumonia and genotypes of Pneumocystis jirovecii in a single tertiary centre in Korea.* BMC Infectious Diseases, 2015. **15**(1): p. 102.

66. Le Gal, S., *et al.*, *AIDS-related Pneumocystis jirovecii genotypes in French Guiana.* Infection, Genetics and Evolution, 2015. **29**: p. 60-67.

67. Friaza, V., *et al.*, *Pneumocystis jiroveci dihydropteroate synthase gene mutations among colonized individuals and Pneumocystis pneumonia patients from Spain.* Postgraduate medicine, 2010. **122**(6): p. 24-28.

68. Suárez, I., *et al.*, *Low prevalence of DHFR and DHPS mutations in Pneumocystis jirovecii strains obtained from a German cohort.* Infection, 2017. **45**(3): p. 341- 347.

69. Nevez, G., *et al.*, *Pneumocystis jirovecii and Cystic Fibrosis in Brittany, France.* Mycopathologia, 2018. **183**(1): p. 81-87.

70. Montesinos, I., *et al.*, *Evaluation of a new commercial real-time PCR assay for diagnosis of Pneumocystis jirovecii pneumonia and identification of dihydropteroate synthase (DHPS) mutations.* Diagnostic microbiology and infectious disease, 2017. **87**(1): p. 32-36.

71. Ricci, G., *et al.*, *Genetic diversity of Pneumocystis jirovecii from a cluster of cases of pneumonia in renal transplant patients: Cross-sectional study.* Mycoses, 2018. **61**(11): p. 845-852.

72. Ozkoc, S., C. Erguden, and S.B. Delibas, *Absence of dihydropteroate synthase gene mutations in Pneumocystis jirovecii strains isolated from Aegean region of Turkey.* Parasitology research, 2018. **117**(10): p. 3103-3108.

73. Szydłowicz, M., *et al.*, *Prevalence and genotyping of Pneumocystis jirovecii in renal transplant recipients—preliminary report.* Parasitology Research, 2019. **118**(1): p. 181-189.

74. Singh, Y., *et al.*, *Genetic polymorphisms associated with treatment failure and mortality in pediatric Pneumocystosis.* Scientific Reports, 2019. **9**(1): p. 1192.

75. Muñoz, C., *et al.*, *Molecular diagnosis and detection of Pneumocystis jirovecii DHPS and DHFR genotypes in respiratory specimens from Colombian patients.* Diagnostic microbiology and infectious disease, 2012. **72**(3): p. 204-213.

76. Taylor, S.M., *et al.*, *Low prevalence of Pneumocystis pneumonia (PCP) but high prevalence of Pneumocystis dihydropteroate synthase (dhps) gene mutations in HIV-infected persons in Uganda.* PLoS One, 2012. **7**(11): p. e49991.

77. Gurbuz, C.E., *et al.*, *Genetic diversity of Pneumocystis jirovecii isolates among Turkish population based on mitochondrial large subunit ribosomal RNA and dihydropteroate synthase gene typing.* Medical Mycology, 2021. **59**(8): p. 813- 820.

78. Tsvetkova, N., *et al. A Pilot Study of Drug Resistance in Pneumocystis Jirovecii Isolates from Bulgarian Patients with Pneumocystis Pneumonia*. in *Conference Proceedings, Istanbul Turkey Jun*. 2021.

79. Nawel, B., *et al.*, *Pneumocystis Jirovecii Pneumonia: Molecular Diagnosis and Antimicrobial Susceptibility in Two Renal Transplants Cases.* J Clin Case Stu, 2021. **6**(2).

80. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.

81. Kumar, S., *et al.*, *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. **35**(6): p. 1547-1549.

82. Mangal, T.D., *et al.*, *Estimating HIV incidence from surveillance data indicates a second wave of infections in Brazil.* Epidemics, 2019. **27**: p. 77-85.

83. Ponce, C.A., *et al.*, *High prevalence of Pneumocystis jirovecii dihydropteroate synthase gene mutations in patients with a first episode of Pneumocystis pneumonia in Santiago, Chile, and clinical response to trimethoprim- sulfamethoxazole therapy.* Antimicrobial agents and chemotherapy, 2017. **61**(2): p. e01290-16.

Chapter 7

Pneumocystis Detection and Typing using Next Generation Sequencing

7.1 Next Generation Sequencing Overview

Multilocus sequence typing has been considered to be the gold standard for genotyping of *Pneumocystis,* and is mainly based on first generation sequencing technology, also referred to as Sanger sequencing [1]. Despite being the gold standard, it is not to say that it does not come without its own limitations, some demonstrated and addressed in previous chapters.The major limitation of Sanger sequencing is the inability to sequence DNA from genetically different strains simultaneously.

Next generation sequencing (NGS) refers to the recent generation of technologies developed to combat pitfalls of previous sequencing technologies, by allowing for, high throughput culture independent sequencing of multiple organisms simultaneously [2]. Since its debut in 2004, the cost of sequencing using NGS technologies has significantly dropped, and it became a regular platform for taxonomic identification [3]. In the last decade, NGS hasbeen applied directly to clinical specimens, for unbiased detection of microbiological pathogens [4-6]. dubbed metagenomic NGS, allowing for broad identification of microbes, which can result in the identification of unsuspected, unknown, or undiscovered pathogens [7]. Opposed to only targeting specific regions within the genome, such as Sanger sequencing, metagenomic based NGS sequences all nucleic acids in a sample, and then compares the obtained sequences across reference genomes to identify which organisms are present and enables the calculations of the numbers of the sequences are present in a sample [3]. The sequencing of all DNA or RNA within a sample allows for an in-depth analysis and has the ability to determine between major or minor (and even reported ultra-minor) microbial populations in a community [8]. Metagenomic based NGS has also been pairedwith targeted sequencing (metabarcoding), such as with the use of primers for high conservedareas among eukaryotes and prokaryotes (16S rRNA or ITS) [9, 10]. A study from Miao *et*

al., showed that for infectious disease diagnosis, using metagenomic based NGS resulted in 50.7% sensitivity and 85.7% specificity, performing significantly better than culture-based diagnosis, especially with regards to fungi, as well as viruses, anaerobes and bacteria, such as *Mycobacterium tuberculosis* [11]. Li *et al.* found that the sensitivity was recorded at 100%, and specificity at 76.5% for bacterial, and 57.1% and 61.5%, respectively, for fungal diagnosis. For bacterial and viral investigations, metagenomic based NGS has been used for distinguishing between strains and helping to establish linages and evolutionary trails of outbreaks, as well as investigate links with antimicrobial resistance, but this has not yet been demonstrated for fungal epidemiological investigations [12-15].

Several sequencing platforms have been developed from various companies, offering a variety based on different methods and appropriate in numerous contexts. Examples ofpopular systems include Illumina from California, USA, that has several platforms (HiSeq, iSeq, MiSeq, NexSeq and NovaSeq), these platforms have the highest throughput on the market, but due to this, also have a high chance of false positives. Also from the US, in Massachusetts, Thermo Fisher Scientific has developed the Ion Torrent and the 5500XL SOLiD, which undertake NGS by applying bead emulsion [16, 17]. From China the Beijing Genomics Institute has developed the BGISEQ-500 platform, which uses combinatorial probe anchor synthesis for sequencing [18]. Finally, Oxford Nanopore Technologies originating from Oxford, UK, offers the first portable sequencing system, which operates in three platforms PromethION, MinION, or GridION [19]. The Nanopore technology uses electrical currents to passes single stranded DNA through a grid of nanopores which are embedded into a polymer membrane, which in turn allows for real time DNA sequencing.The sequencing is obtained by measuring changes within the electrical current created by the nucleotides when passing through the pore. None of the Nanopore platforms require PCR amplification, and the library preparation is relatively fast, involving two major steps, DNA

fragmentation and the ligation of adaptors [3]. The library preparation utilized by this type of shotgun metagenomic based NGS is highly unbiased and allows for the sample to be sequences in terms of prevalence [3]. This technology produces considerably faster results than Illumina sequencing, and is magnitudes faster than more traditional methods, which is very promising for future use in time critical diagnoses of serious infections [3, 5]. Specifically, the MinION from Oxford Nanopore is the smallest sequencing platform on the market, fitting comfortably in the palm of one's hand, able to be plugged into any computer via a simple USB port. The combination of the portable nature, ultra-long read output, quick turn-around time, and cost-effectiveness is perfectly tailored for infectious diseases detection directly from clinical samples [20, 21].

7.2 NGS and *Pneumocystis jirovecii*

Since the introduction of NGS, there have been a handful of publications which have utilised NGS technologies for the detection and whole genome sequencing of. *P. jirovecii* [22, 23]. The first set of publications that explored NGS used pyrosequencing. Pyrosequencing is a type of third generation sequencing method, which amplifies the DNA using PCR, incorporating cycles of dNTPs with the DNA. The DNA nucleotides are then detected in the form of light, which is created when the DNA and the dNTPs bind and release pyrophosphate[24]. In 2012, Cissé *et al.* applied NGS for *de novo* assembly of the *P. jirovecii* genome usingRoche 454 pyrosequencing, which uses emulsion PCR, from a single BAL sample of apatient [25]. In 2016 Alanio *et al.* used the Roche platform GS Junior for pyrosequencing, which revealed that *P. jirovecii* infections are largely mixed infections [22]. When using primers specific to nuclear gene loci 80% of the samples showed mixed sequences, and when using mitochondrial genes this jumped to 92%, a trend which was not detected when using Sanger sequencing. A year later, Charpentier *et al.* used pyrosequencing for genotyping, in a

combination with MLST gene targets, dubbing the technology NGS-MLST [8]. Also using GS Junior by Roche, the study used primer to target the following genetic loci: *CYB, SOD*and *mt26S*, and were able to detect major and minor genotypes within a sample, to confidently map out the transmission of PCP within solid organ transplants in a hospital. Since then, there have been no further analysis of *Pneumocystis* using pyrosequencing. Thisis possible due to the short length of nucleotide sequences produces, which makes the data analysis complex and time consuming to be unreasonable for rapid diagnostics.

It was not until 2019 when two papers discussed using metagenomic based NGS for the identification of pathogens from samples to diagnose infections. Camargo *et al.* used cell-free DNA from the plasma of 10 patients and were able to accurately detect PCP in a patient which had confirmed PCP [26]. The study utilised the registered Karius Test for the NGS, which is a commercial platform and needs to be outsourced for sequencing and analysis. Although promising results from the small sample set were obtained further analysis should be undertaken, as this could be a potential alternative testing for patients who do not tolerate well the collection of respiratory samples. The second paper was a study conducted by Zhang*et al.*, that used shotgun metagenomic based NGS (with no specific platform stated) on 13 patients with pneumonia of unknown etiologic [27]. They successfully detected and diagnosed *P. jirovecii* as the most probable cause of pneumonia, as well as identifying secondary pathogens within the clinical specimens, which were most likely causing co- infections. Only three of the samples were positive for *P. jirovecii* when utilising traditional microscopy. Furthermore, the study also reported the detection *P. jirovecii* within threeperipheral blood samples, further reinforcing the possibility of using non-respiratory samples in future diagnosis. Considering the development and cost decrease of metagenomic based NGS over the past few years, it is crucial to further examine the ability to use third generation metagenomic based NGS technology for the detection of *Pneumocystis* after those promising

preliminary findings. In addition, it would be further worthwhile to explore the possibility of utilising metagenomic based NGS also for genotyping, as done with the previous NGS- MLST attempt.

The most promising NGS technology to date for rapid diagnosis of infectious diseases remains to be the Nanopore MinION. Successful diagnosis of infectious diseases, such as PCP infections which remain to be a struggle for clinicians to swiftly diagnose, usingportable, rapid diagnostics would be a game changer for battling nosocomial PCP outbreaks.

Long-read sequencing based clinical metagenomics for the detection and confirmation of *Pneumocystis jirovecii* directly from clinical specimens: A paradigm shift in mycological diagnostics

This manuscript was published in the Medical Mycology in July 2020 (volume 58, issue 5, pages 650-660), but published online as of the 23rd of November 2019. L. Irinyi is the primary author and contributed to all aspects of this study (Designing and performing the research, analyzing the data, and preparing the manuscript). M. T. V. Hoang was involved in the acquisition and analysis of data, the design of the research and preparing the manuscript. L. Pasic was involved in the acquisition of data, the design of the research and preparing the manuscript. Y. Hu, J. Rathjen, E. Stone and B. Schwessinger were involved in the data analysis. I. Basu, W. McKinney, A. J. Morris were involved in sample acquisition. W. Meyer was involved in designing of the research, project administration and funding acquisition. All authors were involved in the revision of the manuscript and support its inclusion in this thesis.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Lana Pasic

[redacted]

11 October 2021

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Wieland Meyer

[redacted]

11 October 2021

ISHAM INTERNATIONAL SOCIETY FOR HUMAN AND ANIMAL MYCOLOGY

Medical Mycology, 2020, 58, 650–660
doi:10.1093/mmy/myz109
Advance Access Publication Date: 23 November 2019
Original Article

Original Article

Long-read sequencing based clinical metagenomics for the detection and confirmation of *Pneumocystis jirovecii* directly from clinical specimens: A paradigm shift in mycological diagnostics

Laszlo Irinyi[1,2,3,†], Yiheng Hu[6,†], Minh Thuy Vi Hoang[1,2,3], Lana Pasic[1,2,3], Catriona Halliday[4], Menuk Jayawardena[4], Indira Basu[8], Wendy McKinney[8], Arthur J. Morris[8], John Rathjen[6], Eric Stone[6,7], Sharon Chen[1,2,4], Tania C. Sorrell[1,2], Benjamin Schwessinger[6,*] and Wieland Meyer[1,2,3,5,*]

[1]Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, The University of Sydney, Sydney, NSW, Australia, [2]Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney, NSW, Australia, [3]Westmead Institute for Medical Research, Westmead, NSW Australia, [4]Centre for Infectious Diseases and Microbiology-Laboratory Services, Institute for Clinical Pathology and Medical Research, NSW Health Pathology, Westmead, NSW, Australia, [5]Westmead Hospital (Research and Education Network), Westmead, NSW, Australia, [6]Research School of Biology, Australian National University, Canberra, ACT, Australia, [7]ANU-CSIRO Centre for Genomics, Metabolomics and Bioinformatics, Canberra, ACT, Australia and [8]Microbiology Department, LabPLUS, Auckland City Hospital, Auckland, New Zealand

*To whom correspondence should be addressed. Prof. Wieland Meyer, PhD, Molecular Mycology Research Laboratory, CIDM, Westmead Institute for Medical Research, 176. Hawkesbury Road, Westmead, NSW 2145, Australia, Tel: +61 2 86273430, E-mail: wieland.meyer@sydney.edu.au and, Benjamin Schwessinger, PhD, Division of Plant Science, Research School of Biology, College of Science, Linnaeus Building (134), Linnaeus Way The Australian National University, Canberra ACT 0200 Australia, Tel: +61 2 61257794, E-mail: benjamin.schwessinger@anu.edu.au

†L.I., and Y.H. are equally contributed to this manuscript.

Received 18 July 2019; Revised 26 August 2019; Accepted 19 November 2019; Editorial Decision 25 September 2019

Abstract

The advent of next generation sequencing technologies has enabled the characterization of the genetic content of entire communities of organisms, including those in clinical specimens, without prior culturing. The MinION from Oxford Nanopore Technologies offers real-time, direct sequencing of long DNA fragments directly from clinical samples. The aim of this study was to assess the ability of unbiased, genome-wide, long-read, shotgun sequencing using MinION to identify *Pneumocystis jirovecii* directly from respiratory tract specimens and to characterize the associated mycobiome. *Pneumocystis* pneumonia (PCP) is a life-threatening fungal disease caused by *P. jirovecii*. Currently, the diagnosis of PCP relies on direct microscopic or real-time quantitative polymerase chain reaction (PCR) examination of respiratory tract specimens, as *P. jirovecii* cannot be cultured readily *in vitro*. *P. jirovecii* DNA was detected in bronchoalveolar lavage (BAL) and induced sputum (IS) samples from three patients with confirmed PCP. Other fungi present in the associated mycobiome included known human pathogens (*Aspergillus, Cryptococcus, Pichia*) as well as commensal species (*Candida, Malassezia, Bipolaris*). We have established optimized sample preparation conditions for the generation of high-quality data, curated databases, and data analysis tools, which are key to the application of long-read MinION sequencing leading to a fundamental new approach in fungal diagnostics.

Key words: mycoses, molecular diagnostics, clinical metagenomics, MinION, *Pneumocystis*.

Introduction

Rapid, reliable, and accurate detection of fungal pathogens enables early initiation of appropriate antifungal therapy and improved clinical outcomes of invasive fungal diseases. Conventional diagnostic methods have limitations, such as low sensitivity and/or specificity, long turnaround times, are time consuming, and have a restricted ability to detect multiple pathogens simultaneously. Next generation sequencing (NGS) is a powerful, culture-independent tool that can characterize mixed microbial communities.[1] The feasibility of direct fungal identification and characterization by NGS short-read sequencing has been demonstrated in human oral and respiratory mycobiome studies.[2,3] Third generation sequencing technologies, such as nanopore sequencing (Oxford Nanopore Technologies [ONT]), generate long sequence reads (kb to Mb), which enable more accurate mapping to reference genomes and real-time data analysis,[4] with superior discriminatory power in microbial identification.[5] Currently, the smallest sequencing platform offered by ONT is the MinION, a palm-sized sequencing device capable of generating ultra-long sequence reads in real-time, with low initial start-up costs and ease of use. It is uniquely suited for rapid diagnostics. The instrument has been applied successfully in bacterial outbreak investigation, real-time virus detection, and antibiotic resistance studies.[6] However, direct detection of pathogens in clinical specimens using a metagenomics approach is in its infancy and no application has previously been trialed for fungal pathogens.

The fungal microbiome (mycobiome) of bronchoalveolar lavage (BAL) and induced sputum (IS) is not well characterized, particularly in relation to infectious fungal diseases, such as candidiasis or *Pneumocystis* pneumonia (PCP). This is mainly because the respiratory tract was considered to be a sterile, harsh environment with a strong immune surveillance and relatively limited access to nutrients.[7] Until recent years, studies using short-read sequencing described complex microbiomes from the respiratory tract, including smokers and people with different respiratory diseases.[7,8] However, most of those studies only well characterized the bacterial compositions, while the fungal community remains largely unknown.

PCP, due to the opportunistic pathogen *Pneumocystis jirovecii*, has a high morbidity and mortality in patients infected with human immunodeficiency virus (HIV), organ transplant recipients, and patients with other causes of immunocompromise.[9] Diagnosis relies on identification of fungal cysts (or trophozoites) by direct microscopy on IS or BAL fluid using giemsa, toluidine blue or commercially available *Pneumocystis* specific immunofluorescent probes, but identification success varies and is largely dependent on the experience and skill of the laboratory personnel.[10,11] To overcome these issues, various polymerase chain reaction (PCR) based methods have been developed, such as standard or quantitative PCR usually targeting 400–700 bp long fragments.[12–16] Previous studies have shown that the sensitivity of PCR based methods can be improved by using a multicopy gene target, such as the mtLSU rRNA or the major surface glycoprotein,[14,17,18] or to use nested PCR to increase the detection rate.[19]

This study aimed to prove the utility of nanopore based metagenomics as part of the diagnosis of laboratory-confirmed cases of PCP from respiratory samples and to characterize the associated mycobiome of *P. jirovecii*.

Methods

Ethics

Ethical approval was not required for this as per institutional guidelines.

Sample collection and storage

We collected four induced sputum (IS) and two bronchoalveolar lavage (BAL) samples used for routine microbiological examination, including culture and/or *P. jirovecii* PCR. Three clinical samples were PCR positive for *Pneumocystis*, as per previous routine diagnosis (Pj_1, Pj_2, and Pj_3), and the other three were used as controls (C_1, C_2, and C_3) (Table 1).

Routine clinical microbiological testing

P. jirovecii DNA was detected using an "in house" real-time quantitative PCR assay that targets the *β-tubulin* or *surface glycoprotein* gene.[20]

DNA extraction for metagenomics analysis

Total DNA was extracted using the BAL DNA Isolation Kit (cat. no. 46 200, Norgen, Canada) following Irinyi et al.[21] adapted from the manufacturer's instructions. The detailed DNA extraction protocol can be found at: protocol.io (dx.doi.org/10.17504/protocols.io.6enhbde). Briefly, samples were ground in liquid nitrogen to break fungal cell walls before the commercial kit was used. To extract DNA, slurry D solution was added to liquefy the viscous sputum samples and mixed well. After a 5 minute low speed centrifuge step, Proteinase K and Lysozyme were added to the precipitated slurry pellet and incubated at 60°C for 20 minutes. After the incubation, the mixture was subjected to three washing steps with solutions supplied in the commercial Kit and ethanol. The DNA from the spin-column was eluted with 100 µl sterile water. The DNA concentration of the samples was determined using DeNovix ds-DNA Broad Range Fluorescent Assay Kit (Denovix, Wilmington, DE, USA). The average size of DNA was determined by agarose gel electrophoresis (1.5%) of 100 ng DNA for 1 hour at 100 volts.

Table 1. Sample characteristics, summary of nanopore sequencing and taxonomy assignments obtained by using WIMP and BLAST for four induced sputum (IS) (C_1, C_2, C_3, Pj_1) and two bronchoalveolar lavage (BAL) specimens (Pj_2, Pj_3).

Sample	C_1		C_2		C_3		Pj_1		Pj_2		Pj_3	
Origin	CIDMLS, ICPMR, Westmead Hospital IS		CIDMLS, ICPMR, Westmead Hospital IS		CIDMLS, ICPMR, Westmead Hospital IS		Auckland City Hospital IS		CIDMLS, ICPMR, Westmead Hospital BAL		CIDMLS, ICPMR, Westmead Hospital BAL	
Specimen type												
Pneumocystis PCR result	Not Performed		Negative		Not Performed		Positive		Positive		Positive	
Total amount of DNA used in the DNA library preparation (ng)	100		100		100		100		33		33	
Total yield (Mbps)	2208		882		1599		49		388		320	
Mean length (bp)	626		1129		611		657		727		836	
No. of reads	3526,787		781,511		2616,428		81,956		535,411		383,693	
No. of passed QC reads*	3525,062 (99.95%)		781,160 (99.96%)		2615,142 (99.95%)		67,251 (82.06%)		499,125 (93.22%)		359,230 (93.62%)	
Workflow	WIMP	BLAST	WIMP	BLAST	WIMP	BLAST	WIMP	BLAST	WIMP	BLAST	WIMP	BLAST
No. of classified reads	3081,681 (87.42%)	3179,709 (90.2%)	747,688 (95.72%)	753,110 (96.41%)	2268,438 (86.74%)	2325,136 (88.91%)	27,500 (40.89%)	36,055 (53.61%)	467,325 (93.67%)	479,572 (96.08%)	329,200 (91.64%)	338,625 (94.26%)
No. of unclassified reads	443,381 (12.58%)	345,353 (9.8%)	33,472 (4.28%)	28,050 (3.59%)	346,704 (13.26%)	290,006 (11.09%)	39,751 (59.11%)	31,196 (46.39%)	31,600 (6.33%)	19,553 (3.92%)	30,030 (8.36%)	20,605 (5.74%)
No. of human reads (human/mammal)	2921,496 (82.88%)	2917,433/77,385 (82.76%/2.2%)	744,046 (95.25%)	734,343/11,133 (94.01%/1.43%)	2244,710 (85.84%)	2230,428/66,793 (85.29%/2.55%)	7010 (10.42%)	7099/137 (10.5%)/(0.2%)	342,960 (68.71%)	345,145/5956 (69.15%)/(1.19%)	318,312 (88.61%)	317,374/6120 (88.35%)/(1.7%)
No. of microbial reads	160,185 (4.54%)	184,891 (5.25%)	3642 (0.47%)	7634 (0.98%)	23,728 (0.91%)	27,915 (1.07%)	20,490 (30.47%)	28,819 (42.85%)	124,565 (24.96%)	128,471 (25.74%)	10,888 (3.03%)	15,131 (4.21%)
No. of bacterial reads	139,749 (3.96%)	156,531 (4.44%)	872 (0.11%)	1446 (0.19%)	5189 (0.2%)	7971 (0.3%)	20,274 (30.15%)	25,495 (37.91%)	123,195 (24.68%)	126,793 (25.4%)	9939 (2.77%)	12,437 (3.46%)
No. of fungal reads	9583 (0.27%)	73 (0.002%)	1315 (0.17%)	18 (0.0023%)	8847 (0.34%)	77 (0.003%)	124 (0.18%)	57 (0.08%)	589 (0.12%)	69 (0.01%)	532 (0.15%)	6 (0.0016%)
No. of *Pneumocystis* reads	11	0	2	0	19	1	14	14	9	9	1	2

Library preparation

DNA sequencing libraries were prepared using the Ligation Sequencing 1D SQK-LSK108 and Native Barcoding Expansion (PCR-free) EXP-NBD103 Kits from ONT (UK), as adapted by Hu and Schwessinger,[22] which was adapted from the manufacturer's instructions with the omission of DNA fragmentation and DNA repair. DNA was first cleaned up using a 1× volume of Agencourt AMPure XP beads (cat. no. A63881, Beckman Coulter, Indianapolis, IN, USA), incubated at room temperature with gentle mixing for 5 minutes, washed twice with 200 μl fresh 70% ethanol, the pellet was allowed to dry for 2 minutes, and the DNA was eluted in 51 μl nuclease free water and quantified using NanoDrop® (Thermo Fisher Scientific, Waltham, MA, USA) and Promega Quantus™ Fluorometer (cat. No. E6150, Promega, Madison. WI, USA) follow the manufacturer's instructions. All DNA samples showed an absorbance ratio of A260/A280 >1.8 and an A260/A230 ratio of >2.0 from the NanoDrop®. DNA was end-repaired using NEBNext Ultra II End-Repair/dA-tailing Module (cat. No. E7546, New England Biolabs [NEB], Ipswich, MA, USA) by adding 7 μl Ultra II End-Prep buffer, 3 μl Ultra II End-Prep enzyme mix. The mixture was incubated at 20°C for 10 minutes and 65°C for 10 minutes. A 1× volume (60 μl) Agencourt AMPure XP clean-up was performed and the DNA was eluted in 31 μl nuclease free water. Barcoding reaction was performed by adding 2 μl of each native barcode and 20 μl NEB Blunt/TA Master Mix (cat. No. M0367, New England Biolabs [NEB], USA) into 18 μl DNA, mixing gently and incubating at room temperature for 10 minutes. A 1× volume (40 μl) Agencourt AMPure XP clean-up was then performed and the DNA was eluted in 15 μl nuclease free water. Ligation was then performed by adding 20 μl Barcode Adapter Mix (EXP-NBD103 Native Barcoding Expansion Kit, ONT, UK), 20 μl NEBNext Quick Ligation Reaction Buffer, and Quick T4 DNA Ligase (cat. No. E6056, New England Biolabs [(NEB], USA) to the 50 μl pooled equimolar barcoded DNA, mixing gently and incubating at room temperature for 10 minutes. The adapter-ligated DNA was cleaned-up by adding a 0.4× volume (40 μl) of Agencourt AMPure XP beads, incubating for 5 minutes at room temperature and resuspending the pellet twice in 140 μl ABB provided in the SQK-LSK108 kit. The purified-ligated DNA was resuspended by adding 15 μl ELB provided in the SQK-LSK108 (ONT, UK) kit and resuspending the beads. The beads were pelleted again and the supernatant transferred to a new 0.5 ml DNA LoBind tube (cat. no. 00 301 22348, Eppendorf, Germany).

Sequencing

In total, three independent sequencing reactions were performed on a MinION flow cell (R9.4, ONT) connected to a MK1B device (ONT) operated by the MinKNOW software (version 2.0.2): one reaction each for sample Pj_1, the control group and samples Pj_2 and Pj_3. Each flow cell was primed with 1 ml of priming buffer comprising 480 μl Running Buffer Fuel Mix (RBF, ONT) and 520 μl nuclease free water. In total, 12 μl of amplicon library was added to a loading mix including 35 μl RBF, 25.5 μl Library Loading beads (ONT library loading bead kit EXP-LLB001, batch number EB01.10.0012) and 2.5 μl water with a final volume of 75 μl and then added to the flow cell via the SpotON sample port. The 'NC_48Hr_sequencing_FLO-MIN106_SQK-LSK108' protocol was executed through MinKNOW after loading the library, and run for 48 hours. Raw fast5 files were processed using Albacore 2.3.1 software (ONT) for basecalling, barcode de-multiplexing and quality filtering (Phred quality (Q) score of >7) as per the manufacturer's recommendations. Basecalled fastq files were uploaded into NCBI Short Reads Archive under the Bioproject number PRJNA524377 including six accession numbers: SRX5439674, SRX5439673, SRX5439672, SRX5439671, SRX5439670, and SRX5439669.

Data analysis

All reads that passed Albacore quality filtering and recognized multiplexed barcodes were treated in parallel as follows: First, barcode and adapter sequences were trimmed from the ends of reads using Porechop.[23] To identify middle adapter sequences by Porechop, a 95% threshold was set, and reads with middle adapter sequences were discarded. Secondly, seqtk (https://github.com/lh3/seqtk) was used to convert the processed fastq file into fasta format. The 'What's in my pot' (WIMP) quantitative online analysis tool[24] and nucleotide BLAST (blastn, Basic Local Alignment Search Tool)[25] analyses were then performed independently. BLAST searches were performed using the fasta files for each barcode as queries against the National Center for Biotechnology Information (NCBI) nucleotide (nt) database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/, downloaded 09/2018). Individual reads were assigned to a specific DNA sequence based on their best blastn hit (E-value <0.01).

After PoreChop, WIMP uses fastq files directly through EPI2ME Desktop Agent (vers. v2.58). A Q score of ≥7 was set in the beginning of the workflow. Fastq files from each barcode were separated into files with 4000 reads within a folder using fastq-splitter.pl script (http://kirill-kryukov.com/study/tools/fastq-splitter/) before uploading onto the WIMP server through the EPI2ME Desktop Agent. The WIMP analysis output file and quality control csv file for each sample were deposited at Figshare https://doi.org/10.6084/m9.figshare.7772852.v1.

We used Python programming language to summarize relevant information from the sequencing summary files and output files from WIMP and BLAST. To facilitate the reproducibility of our results, all python modules, programming environment, and three scripts used for all analysis and figure plotting were cited and deposited at https://github.com/Yiheng323/clinical_diagnosis_of_pneumocystis.git.

Results

Metagenomic sequencing of BAL and IS samples

DNA extraction from IS and BAL samples yielded various amounts of dsDNA per sample (33 ng from BAL samples and hundreds of nanograms from IS samples). The total amount of DNA we used in the library preparation was 33–100 ng for each sample (Table 1). After the sequencing, we used Albacore software to convert raw sequencing signal to nucleotide sequence. Based on read mean Q scores, Albacore grouped reads into pass (Q score >7) or fail (Q score ≤7). We only proceeded with analysis of sequencing reads with a Q score above 7 and reads with barcode sequences identified by Albacore at both ends (see Table 1). The Pj_1 IS sample yielded only 49 Mbps of total with 67,251 reads passing the quality control. The other two *Pneumocystis* positive BAL samples resulted in 388 and 320 Mbps of total sequence length with 499,125 and 359,230 reads and a Q score >7, respectively. Sequencing of the control sample C_1 generated 2208 Mbps total sequences, representing the highest yield from all samples, followed by 1599 Mbps from sample C_3 and 882 Mbps from sample C_2 (Table 1).

Independent WIMP and BLAST taxonomic classification of *P. jirovecii*

To evaluate different analysis and diagnostic strategies we performed WIMP and BLAST independently on all quality filtered sequence data. We assigned the origin of each read based on a 'winner-takes-all' strategy, which considers the best match of each sequencing read defined by the lowest BLAST E-values or the highest WIMP scores. WIMP and BLAST classified 87–95% of reads in each sample, except for the sample Pj_1 where only 40.89% and 53.61% were classified, respectively.

Reads assigned to human (*Homo sapiens*) (or other mammalian species in case of the BLAST) ranged from 70% to 95% of total reads per sample, except for Pj_1 (~10%) (Table 1). The relative high error rate of nanopore sequencing may lead to misclassifications when comparing reads with reference genomes originated from closely related species. Therefore, we choose to examine the fungal population classified by WIMP and BLAST at genus level to tolerate misclassifications within the same genus. From the WIMP classification, the total fungal sequence length comprised 0.10% to 0.26% of the total sequence length of each group. When separating reads assigned to fungal genomes, the proportion of each genus was consistent (Fig. 1). The genus *Malassezia* ranked the highest in all samples, except sample Pj_1, where *Pneumocystis* reads were ranked the highest. Most of the fungal genera identified by WIMP were shared among all samples, such as *Aspergillus* (3.42%–5.28%), *Phycomyces* (3.80%–7.09%), and *Lobosporangium* (3.23%–6.11%). Unexpectedly, *Pneumocystis* reads were detected in both pneumocystis positive and control samples using the WIMP algorithm with an abundance of total fungal sequence length at 0.49%–14.42% and 0.13% – 0.23%, respectively. Although the abundance of *Pneumocystis* in the positive group was higher than that in the *Pneumocystis* control group, this result suggests that the WIMP classification alone does not provide enough confidence for diagnosis. Compared to WIMP, BLAST analysis against the NCBI nucleotide database classified much fewer reads as fungi (Table 1), comprising slightly over 0.1% in the three *Pneumocystis* positive samples and around 0.001% in all the rest of the samples. *Pneumocystis* reads were mostly detected in the *Pneumocystis* positive group, although one read from sample C_3 was assigned to the *P. jirovecii* genome.

While BLAST analysis resulted in less reads classified into fungi, it provided a more close-to-reality diagnosis than the WIMP analysis. However, it is worth noting that except for sample Pj_3, the total sequence length of *Pneumocystis* reads classified by either WIMP or BLAST never ranked first among all fungal reads in any *Pneumocystis* positive sample.

Improved diagnosis of *P. jirovecii* by cross-checking fungal reads identified by WIMP with BLAST classification

Many of the fungal genera to which reads were assigned were plant-associated fungi. Examples include *Trichoderma*, *Colletotrichum*, and *Endocarpon* from the WIMP classification and *Rhizophagus* from the BLAST classification (Fig. 1). It is probable that these represent false positives, since no previous study reported their presence in human IS or BAL samples. To improve the specificity of *P. jirovecii* identification by WIMP or BLAST, we compared their concordant classification result for each sample at different taxonomic ranks. Interestingly, even at the kingdom level, the percentage of concordant reads was very low (Fig. 2A), around 10% of the fungal sequences were concordant in two samples and 0.15%–0.45% in the rest. To verify that this method of comparison was unbiased, we tested if such little overlap by different analysis methods was specific to fungi. Therefore, we performed the same comparison on bacterial reads between BLAST and WIMP. Unlike the result from the fungal classification, the general proportion of total sequence length of bacteria reads by WIMP was similar to the proportion assigned by BLAST within each sample (Table 1), and the majority of classifications were concordant for each read at kingdom level (Fig. 2B).

To further investigate the accordance of the read assignment we extracted the accordant fungal reads of each sample (Table 2). Both BLAST and WIMP analysis identified the same 24 reads as *P. jirovecii* within the *Pneumocystis* positive group. In contrast, no reads were assigned concordantly to *Pneumocystis* in the control group. Both BLAST and WIMP assigned 86 reads to 26 other fungal genera in addition to *P. jirovecii* including *Malassezia*, *Candida*, and *Cryptococcus* (Table 2). Overall, 98

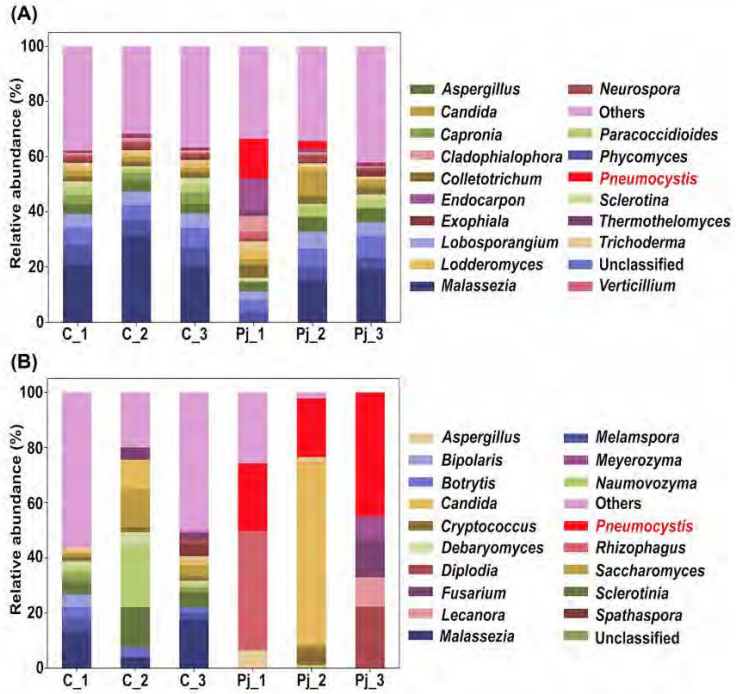


Figure 1. The proportion of total sequence length for 18 most common fungal genera identified by WIMP (**A**) and BLAST (**B**). Columns represent four induced sputum (IS) (C_1, C_2, C_3, Pj_1) and two bronchoalveolar lavage (BAL) specimens (Pj_2, Pj_3). The *y*-axis indicates relative abundance of total sequence lengths for each genus represented by each color. Reads without classified genera but assigned to fungal kingdom were labeled as 'Unclassified.' Fungal genera other than the most common 18 genera were grouped as 'Others.' Genus *Pneumocystis* was labeled with red color.

reads from the concordant 110 fungal reads were assigned to the same species by BLAST and WIMP.

Discussion

The main aim of this study was to apply long-read sequencing to identify human pathogenic fungal species, for example, *P. jirovecii*, directly from clinical samples, such as BAL and IS. We aimed to test if the results obtained by routine diagnostic methods can be confirmed by direct metagenomic shotgun long-read sequencing in order to evaluate its potential as future diagnostic tool for human fungal pathogens. As a result of this study, we propose a nanopore based metagenomics workflow to characterize human fungal pathogens from sample collection to data interpretation (Fig. 3).

MinION based long-read sequencing has been used successfully in a clinical setting to identify Chikungunya, Ebola, and Hepatitis C virus from human blood samples[26] or *Mycobacterium tuberculosis* from direct respiratory samples.[27] In this study, we extended its application to identify fungal pathogens from human respiratory samples, such as *P. jirovecii*, to verify the traditional diagnosis of PCP from a clinical specimen and explored the potential mycobiome of this pathogen.

In addition to *P. jirovecii*, many fungal taxa were detected in the herein studied nonsterile *P. jirovecii* positive BAL and IS specimens. Both the BLAST and WIMP algorithms classified reads into fungal taxa other than *Pneumocystis*, indicating that the disease-causing organism might not necessarily be the most abundant mycobiome member. Without adequate training and additional patient-related disease background data this may lead to the wrong identification of the disease-causing agent. Many of these taxa are part of the human respiratory mycobiome in healthy individuals,[3] such as previously described *Aspergillus* spp., *Candida* spp., *Lobosporangium* spp., and *Malassezia* spp.[28] *Malassezia* spp. are the main member of the microbiome of healthy skin,[29] and their presence in our experiment are likely due to contamination during the sampling process.

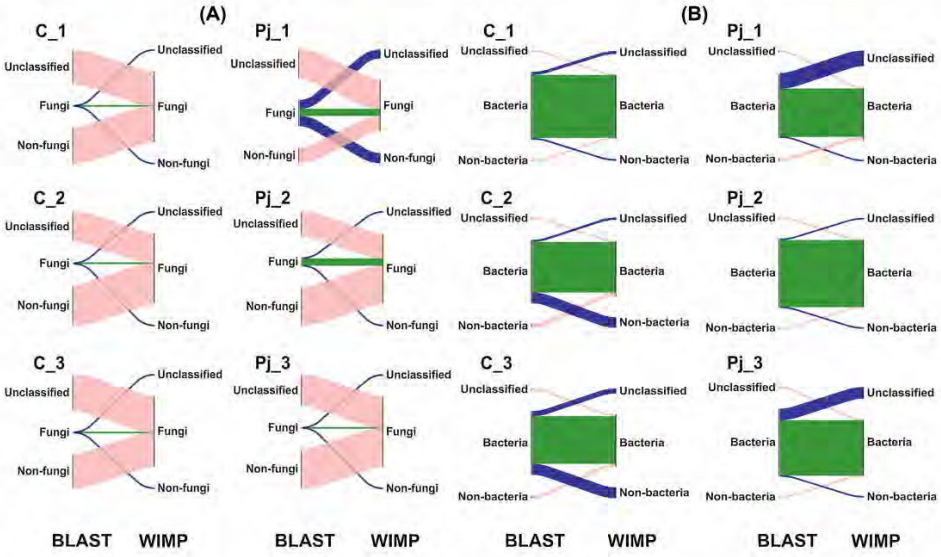


Figure 2. Comparison of fungal reads (**A**) and bacterial reads (**B**) classified by WIMP and BLAST from four induced sputum (IS) (C_1, C_2, C_3, Pj_1) and two bronchoalveolar lavage (BAL) specimens (Pj_2, Pj_3). Green lines indicate species detection concordance, showing a much higher precision for bacteria than fungi, pink lines indicate WIMP and blue lines BLAST specific identifications.

However, some of the genera, such as *Paracoccidioides* spp. identified by the WIMP algorithm herein, are clearly false positives. Species in the genus *Paracoccidioides* can cause systemic mycoses but they are geographically restricted to subtropical areas of Latin America.[30] Therefore, the initial detection of *Paracoccidioides* spp. by the WIMP workflow in our study was alarming. *Paracoccidioides* reads identified by WIMP were then subjected to BLAST search against the NCBI database and none of them were identified as *Paracoccidioides* spp. After crosschecking the nonconcordant reads, such as those corresponding to *Paracoccidioides* spp. and other plant pathogenic fungal species classified by WIMP, many were unclassified even at kingdom level when using BLAST. In the *Pneumocystis* positive group, all *Pneumocystis* reads along other human pathogens or commensal species, such as *Candida* and *Cryptococcus* were concordantly classified by both WIMP and BLAST. We compared all fungal, Chordata, and bacterial reads classified by WIMP and BLAST to see whether they are concordant. We found that bacterial and Chordata reads are largely concordant, while the fungal reads are not. It is not clear why the level of concordance in the classification of bacteria and Chordata was high but in the classification of fungi was low. One of the reasons might be the low number of fungal species present in the genome database of NCBI compared to bacteria. These results indicated that false-positives and error-prone reads represent a challenge for metagenomics studies, especially at the species level.[31,32] Taxonomic profiling and binning programs proved to be accurate at high taxonomic ranks, but their performance decreased significantly below family level.[32] Another reason might be that fungal genomes are usually poorly annotated and more prone to cross-species genome contamination due to the complex structure of eukaryotic genomes, which may result in incorrect taxonomic binning.[31–35]

The results show that the BLAST analysis delivers a more accurate classification than WIMP. BLAST is the most widely used alignment-based sequence similarity search algorithm.[36] It consists of aligning a query sequence to a sequence database to find the best statistically significant matches, coming with major disadvantages, being generally memory and time-consuming, limiting its use for metagenome-scale sequence data analysis. To overcome these shortcomings, alignment free algorithms have been suggested.[37] In recent years, k-mer-based alignment free metagenomics analysis and comparison methods, for example, Centrifuge (used by WIMP)[35] and Kraken[38] have become popular, providing much faster results.

In our case, the combination of BLAST and WIMP analysis provides a strong confirmation of the *P. jirovecii* read assignment in the *Pneumocystis* positive group and overcomes the false positive classification of *Pneumocystis* in the control group, making a successful diagnosis of PCP feasible. Based on this finding, we suggest a combination of the two approaches in a clinical setting when looking at fungal identification. This enables in the first instance a fast and in the second instance accurate identification,

Table 2. Number of reads classified at species level as fungi by WIMP and BLAST and their concordance in four induced sputum (IS) (C_1, C_2, C_3, Pj_1) and two bronchoalveolar lavage (BAL) specimens (Pj_2, Pj_3).

	C_1			C_2				C_3			
Species	WIMP	BLAST	Concordance	Species	WIMP	BLAST	Concordance	Species	WIMP	BLAST	Concordance
Agaricus bisporus	1	1	1	*Candida albicans*	1	1	1	*Agaricus bisporus*	1	1	1
Aspergillus oryzae	0	1	0	*Malassezia globosa*	2	1	1	*Arthrobotrys oligospora*	1	1	1
Bipolaris maydis	1	1	1	*Pichia kudriavzevii*	1	1	1	*Ascoidea rubescens*	1	1	1
Cryptococcus amylolentus	0	1	0	*Saccharomyces cerevisiae*	1	1	1	*Candida albicans*	0	1	0
Cryptococcus neoformans	0	1	0	*Sclerotinia sclerotiorum*	0	1	0	*Candida tropicalis*	1	0	0
Exserohilum turcicum	1	0	0					*Eremothecium gossypii*	1	1	1
Malassezia globose	4	3	3					*Exophiala mesophila*	1	1	1
Paracoccidioides lutzii	0	1	0					*Kluyveromyces marxianus*	0	1	0
Penicillium rubens	0	1	0					*Lodderomyces elongisporus*	1	0	0
Pseudocercospora fijiensis	0	1	0					*Malassezia globosa*	5	5	4
Sclerotinia sclerotiorum	1	1	1					*Neurospora crassa*	0	1	0
Sugiyamaella lignohabitans	1	0	0					*Sclerotinia sclerotiorum*	1	0	0
Tsuchiyaea wingfieldii	1	0	0					*Spathaspora passalidarum*	1	1	1
Unclassified Fungi	1	0	0					*Trichoderma reesei*	1	1	1
Vanderwaltozyma polyspora	1	0	0					*Trichosporon asahii*	0	1	0
								Unclassified Fungi	1	0	0
Total	12	12	6	Total	5	5	4	Total	16	16	10
Concordance %		50		Concordance %		80		Concordance %		62.5	
	Pj_1			Pj_2				Pj_3			
Species	WIMP	BLAST	Concordance	Species	WIMP	BLAST	Concordance	Species	WIMP	BLAST	Concordance
Coniophora puteana	1	1	1	*Candida auris*	1	1	1	*Meyerozyma guilliermondii*	1	1	1
Leptosphaeria biglobosa	0	1	0	*Candida dubliniensis*	45	45	45	*Pneumocystis jirovecii*	1	1	1
Pneumocystis jirovecii	14	14	14	*Cryptococcus neoformans*	5	5	5				
Thielavia terrestris	1	1	1	*Pneumocystis jirovecii*	9	9	9				
Unclassified Fungi	1	0	0								
Total	17	17	16	Total	60	60	60	Total	2	2	2
Concordance %		94		Concordance %		100		Concordance %		100	

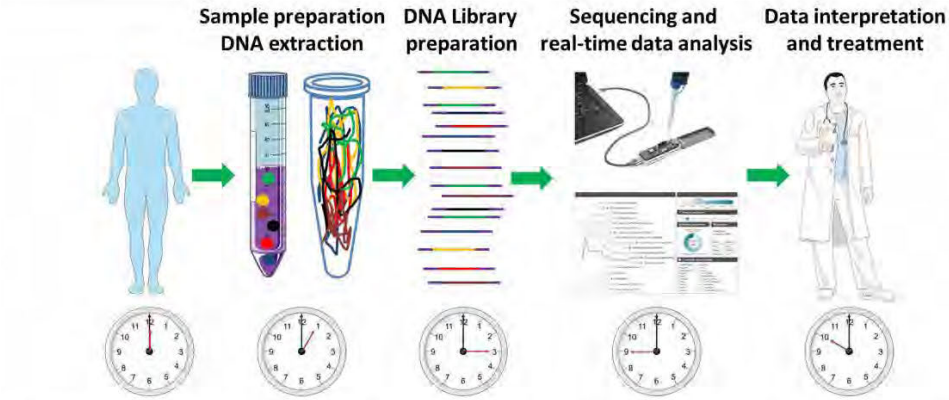


Figure 3. Schematic presentation of the nanopore based metagenomics pipeline to detect and characterize of human pathogens from sample collection to data interpretation.

by running first WIMP to filter out the human reads followed by BLAST to accurately classify the microbial reads.

Another major hurdle to overcome is the low proportion of pathogen DNA in comparison to the host DNA in the sample. This low pathogen DNA content makes it difficult to generate enough target species reads for downstream analyses. Our study is in agreement with previous studies showing that ~95% of raw NGS reads were of human DNA origin.[39] While there are bioinformatic tools to filter out human sequences from a large NGS data set, it still consumed most of the sequencing resources, lowering the sequencing depth of pathogens and the level of cost-effectiveness for pathogen diagnosis. Solutions to this problem can be to specifically lyse human cells and remove the human DNA prior to sequencing,[27,40] or to use microbial DNA enrichment methods to improve the pathogen:host DNA ratio. Both methods have been successfully employed in many studies to enrich bacterial DNA from clinical samples improving the pathogen:host read ratios.[41,42] However, in fungi previous enrichment studies focused on improving the sensitivity of PCR based diagnostics with limited success, and that's why this has rarely been applied to metagenomics studies.[41,43–45]

The success of any DNA based detection method depends on the lysis of the fungal cells in the clinical specimen. Fungi have thick cell walls that impede lysis and the recovery of nucleic acids, such as the *Pneumocystis* cystic form with hard-to-break-up cell walls.[46] Fungal DNA extraction protocols commonly require physical force to rupture the thick cell wall, which has the risk to fragment the DNA. It is crucially important to extract DNA of adequate quality, with a high purity, high molecular weight (HMW), and high concentration, to maximize detection sensitivity and specificity of long-read metagenomics. Therefore, the development of better DNA extraction methods for fungal DNA from complex tissues is a very important research question in regards to improving the accuracy of DNA based fungal pathogen detection.

Another bottleneck of metagenomic shotgun long-read sequencing is the lack of high-quality, curated reference databases of fungal whole genomes. Those databases need to be curated and highly accurate in terms of sequences, metadata, and phylogenetic coverage. False positives and irrelevant organisms, such as *Paracoccidioides* spp. detected in this study are partially due to the incorrect taxonomic identification and annotation of sequences present in the databases.[47]

In conclusion, we provided a proof-of-principle that metagenomic shotgun long-read sequencing with nanopore can identify the presence of a specific disease causing human pathogenic fungus, in our case *P. jirovecii*, directly from a clinical specimen. Yet many challenges (DNA quality, low DNA pathogen:host ratio, lack of quality-controlled reference databases, and bioinformatics tools) are yet to be overcome before a routine application in clinical settings. Despite all the challenges, we anticipate that within the next decade, shotgun metagenomics-based detection and characterization of fungal pathogens will be implemented and translated into routine diagnostic in public health laboratories.

Study limitations

The small sample size in this study is a limitation to detect any specific significant pathogen association or correlation with diseases status; therefore, strong conclusions in identifying pathogen associated sequence signals, other than the confirmation of the causative disease agent, cannot currently be drawn. The sole aim of this study was to apply long-read sequencing to identify fungal species directly from BAL and sputum induced samples.

Acknowledgments

The authors acknowledge the Sydney Informatics Hub and the University of Sydney's high-performance computing cluster Artemis for providing the high-performance computing resources that have contributed to the research results reported within this paper. We thank Joey Lai Genomics Facility manager at the Westmead Institute for Medical Research for his assistance in DNA quantification. We thank Mari Miyamoto from Oxford Nanopore for her technical advice and interest in the project.

This study was supported by a National Health and Medical Research Council of Australia (NH&MRC) grant [no. APP1121936] to W.M., S.C., T.C.S., and Western Sydney Local Health District Research & Education Network Research Grant Scheme to W.M., T.C.S., S.C., and L.I. B.S. is supported by an Australian Research Council Future Fellowship FT180100024, and Y.H., E.S., J.R., and B.S. are supported by The Hermon Slade Foundation grant HSF_17_04. T.C.S. is a Sydney Medical Foundation Fellow.

Declaration of interest

The authors report no conflicts of interest associated with the manuscript. The authors alone are responsible for the content and the writing of the paper.

References

1. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014; 30: 418–426.
2. Ghannoum MA, Jurevic RJ, Mukherjee PK et al. Characterization of the oral fungal microbiome (Mycobiome) in healthy individuals. *PLoS Pathog*. 2010; 6: e1000713.
3. Nguyen LDN, Viscogliosi E, Delhaes L. The lung mycobiome: an emerging field of the human respiratory microbiome. *Front Microbiol*. 2015; 6: 89.
4. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med*. 2016; 8: 97.
5. Benitez-Paez A, Sanz Y. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION portable nanopore sequencer. *GigaScience*. 2017; 6: 1–12.
6. Leggett RM, Clark MD. A world of opportunities with nanopore sequencing. *J Exp Bot*. 2017; 68: 5419–5429.
7. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell*. 2014; 157: 121–141.
8. Salisbury ML, Han MK, Dickson RP, Molyneaux PL. Microbiome in interstitial lung disease: from pathogenesis to treatment target. *Curr Opin Pulm Med*. 2017; 23: 404–410.
9. Maini R, Henderson KL, Sheridan EA et al. Increasing *Pneumocystis* pneumonia, England, UK, 2000–2010. *Emerg Infect Dis*. 2013; 19: 386–392.
10. Sun T, Chess Q, Tanenbaum B. Morphologic criteria for the identification of *Pneumocystis carinii* in Papanicolaou-stained preparations. *Acta Cytol*. 1986; 30: 80–82.
11. Procop GW, Haddad S, Quinn J et al. Detection of *Pneumocystis jiroveci* in respiratory specimens by four staining methods. *J Clin Microbiol*. 2004; 42: 3333–3335.
12. Larsen HH, Masur H, Kovacs JA et al. Development and evaluation of a quantitative, touch-down, real-time PCR assay for diagnosing *Pneumocystis carinii* pneumonia. *J Clin Microbiol*. 2002; 40: 490–494.
13. Robert-Gangneux F, Belaz S, Revest M et al. Diagnosis of *Pneumocystis jirovecii* pneumonia in immunocompromised patients by real-time PCR: a 4-year prospective study. *J Clin Microbiol*. 2014; 52: 3370–3376.
14. Wakefield AE, Pixley FJ, Banerji S et al. Detection of *Pneumocystis carinii* with DNA amplification. *Lancet*. 1990; 336: 451–453.
15. Robberts FJ, Liebowitz LD, Chalkley LJ. Polymerase chain reaction detection of *Pneumocystis jiroveci*: evaluation of 9 assays. *Diagn Microbiol Infect Dis*. 2007; 58: 385–392.
16. Fauchier T, Hasseine L, Gari-Toussaint M, Casanova V, Marty PM, Pomares C. Detection of *Pneumocystis jirovecii* by quantitative PCR to differentiate colonization and pneumonia in immunocompromised HIV-positive and HIV-negative patients. *J Clin Microbiol*. 2016; 54: 1487–1495.
17. Wakefield AE, Pixley FJ, Banerji S et al. Amplification of mitochondrial ribosomal RNA sequences from *Pneumocystis carinii* DNA of rat and human origin. *Mol Biochem Parasitol*. 1990; 43: 69–76.
18. Larsen HH, Huang L, Kovacs JA et al. A prospective, blinded study of quantitative touch-down polymerase chain reaction using oral-wash samples for diagnosis of *Pneumocystis* pneumonia in HIV-infected patients. *J Infect Di*. 2004; 189: 1679–1683.
19. Alvarez-Martinez MJ, Miro JM, Valls ME et al. Sensitivity and specificity of nested and real-time PCR for the detection of *Pneumocystis jiroveci* in clinical specimens. *Diagn Microbiol Infect Dis*. 2006; 56: 153–160.
20. Brancart F, Rodriguez-Villalobos H, Fonteyne PA, Peres-Bota D, Liesnard C. Quantitative TaqMan PCR for detection of *Pneumocystis jiroveci*. *J Microbiol Methods*. 2005; 61: 381–387.
21. Irinyi L, Pasic L, Meyer W. DNA extraction from sputum. 2019; https://www.protocols.io/view/dna-extraction-from-sputum-6enhbde.
22. Hu Y, Schwessinger B. Amplicon sequencing using MinION optimized from 1D native barcoding genomic DNA. 2018; 1–17. https://www.protocols.io/view/amplicon-sequencing-using-minion-optimized-from-1d-mhkc34w.
23. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*. 2017; 3: e000132–e000132.
24. Juul S, Izquierdo F, Hurst A et al. What's in my pot? Real-time species identification on the MinION. *bioRxiv*. 2015: https://doi.org/10.1101/030742.
25. Camacho C, Coulouris G, Avagyan V et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10: 421.
26. Greninger AL, Naccache SN, Federman S et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med*. 2015; 7: 1–13.
27. Charalampous T, Kay GL, Richardson H et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol*. 2019; 37: 783–792.
28. Hogan DA, Willger SD, Dolben EL et al. Analysis of lung microbiota in bronchoalveolar lavage, protected brush and sputum samples from subjects with mild-to-moderate cystic fibrosis lung disease. *PLoS One*. 2016; 11: e0149998.
29. Gaitanis G, Magiatis P, Hantschke M, Bassukas ID, Velegraki A. The *Malassezia* genus in skin and systemic diseases. *Clin Microbiol Rev*. 2012; 25: 106–141.
30. San-Blas G, Nino-Vega G, Iturriaga T. *Paracoccidioides brasiliensis* and paracoccidioidomycosis: molecular approaches to morphogenesis, diagnosis, epidemiology, taxonomy and genetics. *Med Mycol*. 2002; 40: 225–242.
31. Lindgreen S, Adair KL, Gardner PP. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep*. 2016; 6: 19233.
32. Sczyrba A, Hofmann P, Belmann P et al. Critical assessment of metagenome interpretation-a benchmark of metagenomics software. *Nat Methods*. 2017; 14: 1063–1071.
33. Merchant S, Wood DE, Salzberg SL. Unexpected cross-species contamination in genome sequencing projects. *PeerJ*. 2014; 2: e675.
34. Dadi TH, Renard BY, Wieler LH, Semmler T, Reinert K. SLIMM: species level identification of microorganisms from metagenomes. *PeerJ*. 2017; 5: e3138.
35. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 2016; 26: 1721–1729.
36. Boratyn GM, Camacho C, Cooper PS et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res*. 2013; 41: W29–33.
37. Vinga S, Almeida J. Alignment-free sequence comparison-a review. *Bioinformatics*. 2003; 19: 513–523.
38. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014; 15: R46.
39. Bates M, O'Grady J, Maeurer M et al. Assessment of the Xpert MTB/RIF assay for diagnosis of tuberculosis with gastric lavage aspirates in children in sub-Saharan Africa: a prospective descriptive study. *Lancet Infect Dis*. 2013; 13: 36–42.
40. Hasan MR, Rawat A, Tang P et al. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J Clin Microbiol*. 2016; 54: 919–927.

41. Feehery GR, Yigit E, Oyola SO et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One*. 2013; 8: e76096.
42. Loonen AJ, Jansz AR, Stalpers J, Wolffs PF, van den Brule AJ. An evaluation of three processing methods and the effect of reduced culture times for faster direct identification of pathogens from BacT/ALERT blood cultures by MALDI-TOF MS. *Eur J Clin Microbiol Infect Dis*. 2012; 31: 1575–1583.
43. Amorim-Vaz S, Sanglard D. Novel approaches for fungal transcriptomics from host samples. *Front Microbiol*. 2016; 6: 1571–1571.
44. Springer J, Loeffler J, Heinz W et al. Pathogen-specific DNA enrichment does not increase sensitivity of PCR for diagnosis of invasive aspergillosis in neutropenic patients. *J Clin Microbiol*. 2011; 49: 1267–1273.
45. Thoendel M, Jeraldo PR, Greenwood-Quaintance KE et al. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods*. 2016; 127: 141–145.
46. De Stefano JA, Cushion MT, Sleight RG, Walzer PD. Analysis of *Pneumocystis carinii* cyst wall. I. Evidence for an outer surface membrane. *J Protozool*. 1990; 37: 428–435.
47. Bidartondo MI. Preserving accuracy in GenBank. *Science*. 2008; 319: 1616.

7.4 NGS applied to *Pneumocystis canis* infections

7.4.1 *Pneumocystis canis* Background

Pneumocystis species are known to cause the life-threatening fungal disease *Pneumocystis* pneumonia (PCP) in a variety of mammals outside humans, including rabbits, mice, rats, and dogs [28]. *Pneumocystis carinii* f.sp.*'canis' (P. canis*) is the fungal pathogen able to cause infection in canines. Documented as early as the 1970's, PCP has been shown to be life threatening for dogs and is world-wide known to disproportionately affect Cavalier King Charles Spaniels (CKCS), followed by Miniature Dachshunds, as well as reports of infections within Pomeranians in Japan [29-31]. It is theorised that CKCS, and potentially Dachshunds, have an inherited immunodeficiency which heightens their chance of PCP infection, while *P. canis* may live as a commensal fungus within the respiratory tract of other breeds [32]. Just as seen within human infections, it is hypothesised that canine PCP transmission is via airborne droplets, and colonised dogs may play a significant role in acting as a reservoir [31, 33]. Diagnosis is difficult for veterinaries, as diagnosis requires traditional microscopicvisualisation of *Pneumocystis* cysts or trophozoites, which is made difficult as even deep bronchial washings may contain only very little amounts able to be detected [31]. This is made harder as some dogs react poorly to anaesthesia limiting the amount and quality of respiratory sample being able to be collected [34]. Symptoms include dyspnoea, tachypnoea, increased breath, but surprisingly, a cough is often missing, delaying diagnosis and medical care [35, 36]. More often PCP is assumed based of breed characteristic and symptoms, asPCP rarely shows changes in any biochemical tests, blood tests or radiographs [37]. Ironically, timely diagnosis is crucial for recovery, and recovered dogs may need lifelong medication to prevent subsequent infections. In 2017, Danesi *et al.* developed primers to successfully sequence the *mt26S* genetic locus of *P. canis* using nested PCR and qPCR [31].

The study successfully diagnosed PCP in 81% of the suspected-PCP cohort, as well as detecting four colonised *Pneumocystis* cases within the control, detection rates were significantly higher than using traditional microscopy methods. Like to what is demonstrated in *P. jirovecii* diagnosis, molecular methods proved to be a more sensitive and specific method for diagnosing *P. canis*, as well as confidently using qPCR Ct values for differentiating between infections and colonisation.

Molecular methods have shown to effectively diagnose PCP, but the preparation for nested PCR requires a sterile wet laboratory for DNA extraction (if automated extraction is unavailable) and amplification. Since Irinyi *et al.* demonstrated the successful identification of *P. jirovecii* from human PCP patients (see previous section), it was suggested that the Oxford Nanopore MinION sequencer would also able to detect *P. canis* in canine PCP samples. Rapid diagnosis directly from clinical samples when an infection is suspected wouldallow for higher rates of successful treatment and survival of the infected dogs. Additionally, it is important to know if the dog has additional co-infections, such as *Bordetella Bronchiseptica*, which can be an important bacterial PCP co-infection, which needs addressing [34]. The MinION will also be able to quantify other pathogens within the sample,better equipping the clinician with an accurate infection treatment plan.

Two samples from dogs positively diagnosed with PCP were sent for genotyping to our laboratory, enabling us to study the remaining BAL specimen on the Nanopore MinION. The aim was to access the ability of the MinION to successfully identify *P. canis* in PCP positive dogs.

7.4.2 Materials and Methods

7.4.2.1 Canine Patients

Two separate dog BAL samples were sent to the University of Sydney, Faculty of Veterinary Science, Veterinary Pathology Diagnostic Services in November of 2018. Originally, the samples were received at the diagnostic service in May and September 2015, to undergo testing of suspected pneumonia infections. The two samples came from two dogs, Sophie Finney, a 2-year-old Cavalier King Charlies Spaniel, and Prince Simons, a 7-year-old Miniature Fox Terrier. Sophie came in with a suspected pneumocystis infection due to a recurrent cough, and elevated suspicion due to upward of 80 percent of *P. canis* cases arising in the Cavalier King Charlies Spaniel breed. Prince was admitted with savage diffuse bronchointerstitial disease, with the suspicion of possible interstitial fibrosis, atypical pneumonia or Paraquat poisoning, which were then all ruled out when *P. canis* was confirmed.

7.4.2.2 Diagnosis

For diagnosis, both pathologies produced BAL fluid samples from both dogs, with Prince producing moderate amounts, while Sophie produced less. Prince's sample showed only low numbers of inflammatory cells present, which were a mixture of neutrophils and macrophages. Scattered throughout the mucus were several ~5 µm thin-walled spherical or ovoid cysts that contained variable numbers of crescent-shaped trophozoites that possessed small dark nuclear structures, and aggregates of ill-defined grey trophozoites were also numerous external to the cysts. This inflammation along with numerous trophozoites led to the diagnosis of *P. canis* due to its typical characteristics. Sophie's sample had slight blood contamination, so direct smears were made from the mucus and a cytocentrifuged smear was

prepared from the blood contaminated sample. Similarly, like Prince's sample, only low numbers of inflammatory cells were present, which were also a mixture of neutrophils and macrophages. Scattered throughout the mucus were several ~5 μm thin-walled spherical cysts with numerous trophozoites, which was indicative of *P. canis.*

7.4.2.3 NGS Sequencing

The same methodology for DNA preparation, sequencing and data analysis was used as those published in section 7.3. Within the analysis, Dog 1 refers to Prince, and Dog 2 refers to Sophie.

7.5 Results

Both WIMP and BLAST analysis was performed independently on all quality filtered sequence reads. WIMP was able to successfully classify approximately 53% of overall reads identified for the samples of Dog 1 while BLAST classified almost 70%, and lower levels were demonstrated for the samples of Dog 2, 22% of the overall reads classified usingWIMP, and 48% using BLAST (Figure 7.1). WIMP did not detect any dog specific reads, opposed to BLAST which was able to detect approximately 40% in the sample of Dog 1, and 35% in the sample of Dog 2, composing the majority classified reads for both samples.WIMP detected significantly higher rates of dog reads when compares to BLAST, with 45% vs 10% for the sample of Dog 1, and 23% vs 5% for the sample of Dog 2. Conversely, WIMPclassified significantly higher levels of microbial specific reads than BLAST, especially within the sample of Dog 1 samples.

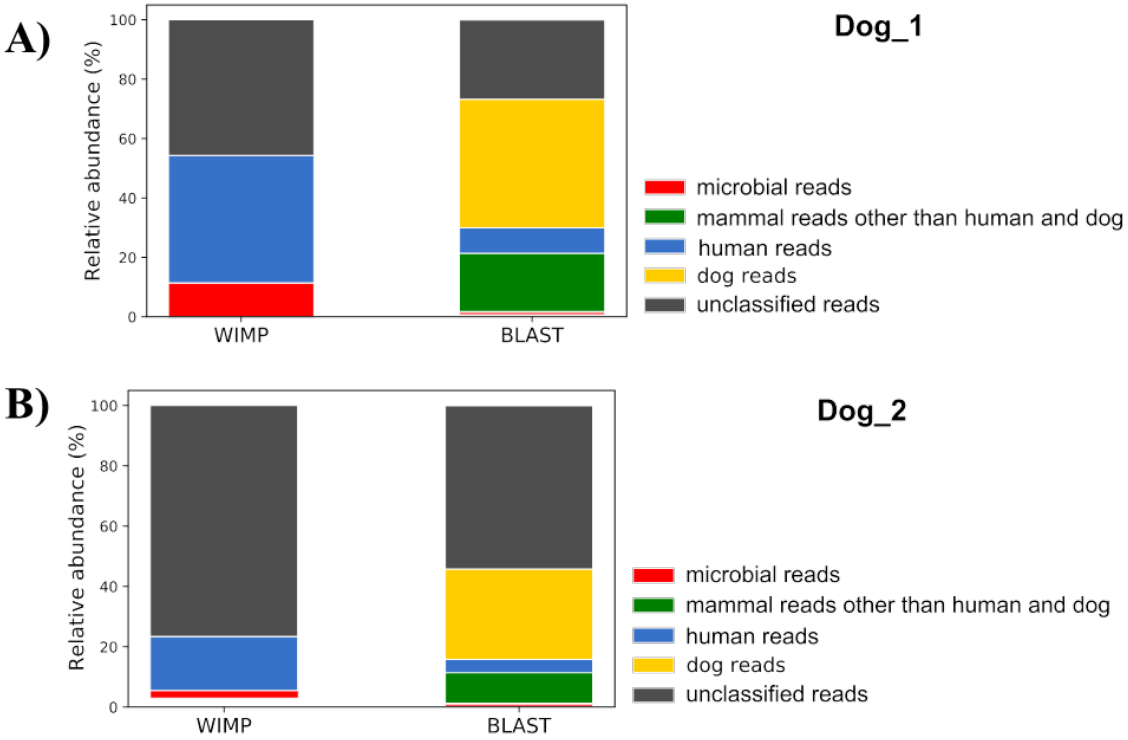
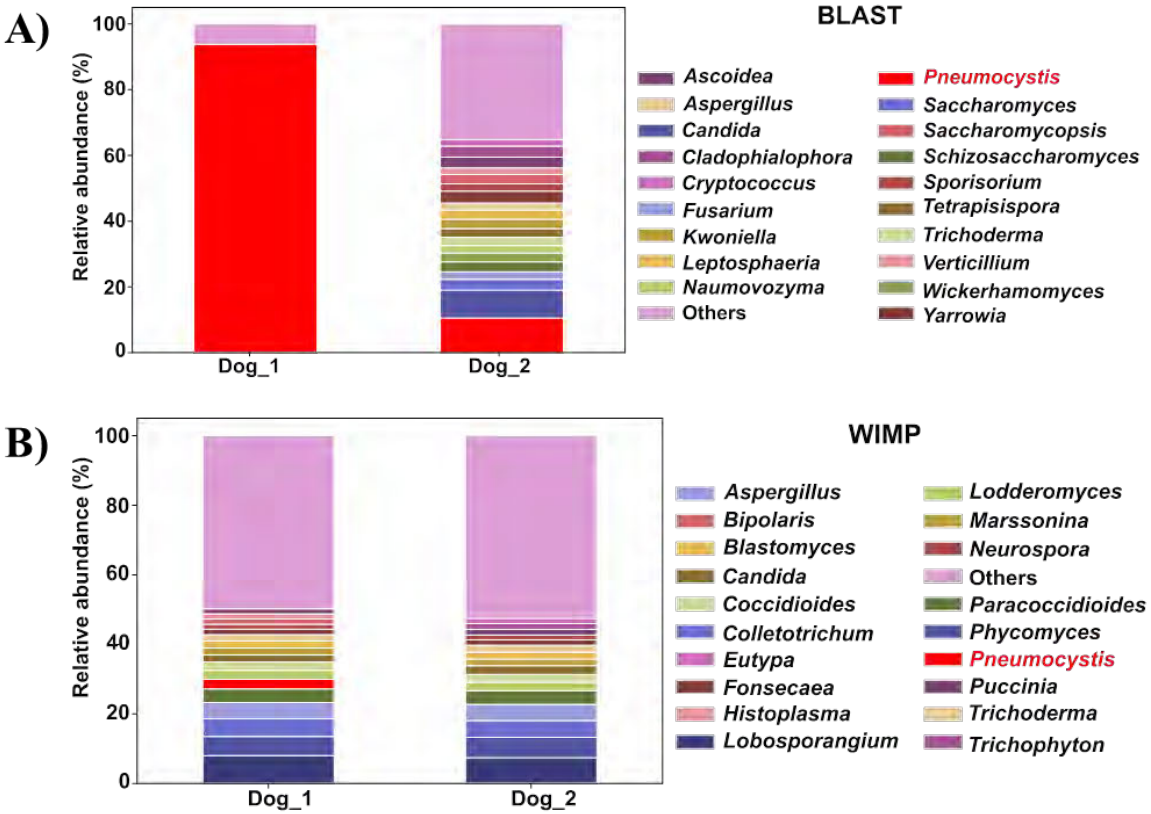


Figure 7.1. Classification of reads classified by WIMP and BLAST from in two dog BAL fluid samples. A) shows classification of Prince coded as Dog 1, and B) shows classification of reads from Sophie. The columns represent the two different reference platforms. The y-axis demonstrates relative abundance of total reads from the two samples.

Due to the relative high error rate of nanopore sequencing, especially when lacking a reference genome, it was decided that the fungal populations would be classified to the genus level, to lower the risk of misclassification at species level. Unsurprisingly, BLAST was able to identify *Pneumocystis* within both samples while WIMP only managed to detect minute levels of *Pneumocystis* in the sample of Dog 1 but did not detect any *Pneumocystis* reads within the sample of Dog 2 (Figure 7.2). Using BLAST, most fungal sequences were classified as *Pneumocystis*, in the sample of Dog 1, with only 10% of reads classified as "Other", signifying fungi which could not be specified at a genus level. Interestingly, noother identifiable fungi were detected. Within the sample of Dog 2, close to 40% of fungi classified as "other" associated with the detection of an additional 19 fungal genera. Despite this, *Pneumocystis* still yielded the highest amount of classified fungal reads, at approximately 10%. *Candida* was the second dominant fungal genus identified, at

approximately 7-8%, while the other genera were identified in minute amounts. In comparison, WIMP over 50% of the reads identified as fungi were unable to be classified at agenus level for both samples. Within both samples, the genera with the most identified reads were *Lobosporangium*, followed by *Phycomyces*, *Colletotrichum*, *Aspergillus* and *Paracoccidioides*, which excluding *Aspergillus*, were not detected within the BLAST analysis.

Figure 7.2. Proportion of total fugal genera identified by A) BLAST, and B) WIMP. The two columns represent



the results from each dog, while the y-axis demonstrates relative abundance of total fungal genera identification from the two samples. Reads which were unclassified to the genus level but identified within the fungal kingdom are labelled as “Others”.

To strengthen the specificity of fungal identification and the accordance of the reads, the fungal reads were compared between the two platforms, and then against all bacterial reads (Figure 7.3). Even at the kingdom level, the percentage of concordant fungal reads was very low in both dogs. Between the dogs, the levels of classified and unclassified reads varied significancy, but both algorithms detected large proportions on non-fungal reads. In

comparison to bacteria, the general proportion of concurrency was much higher within the bacterial reads.

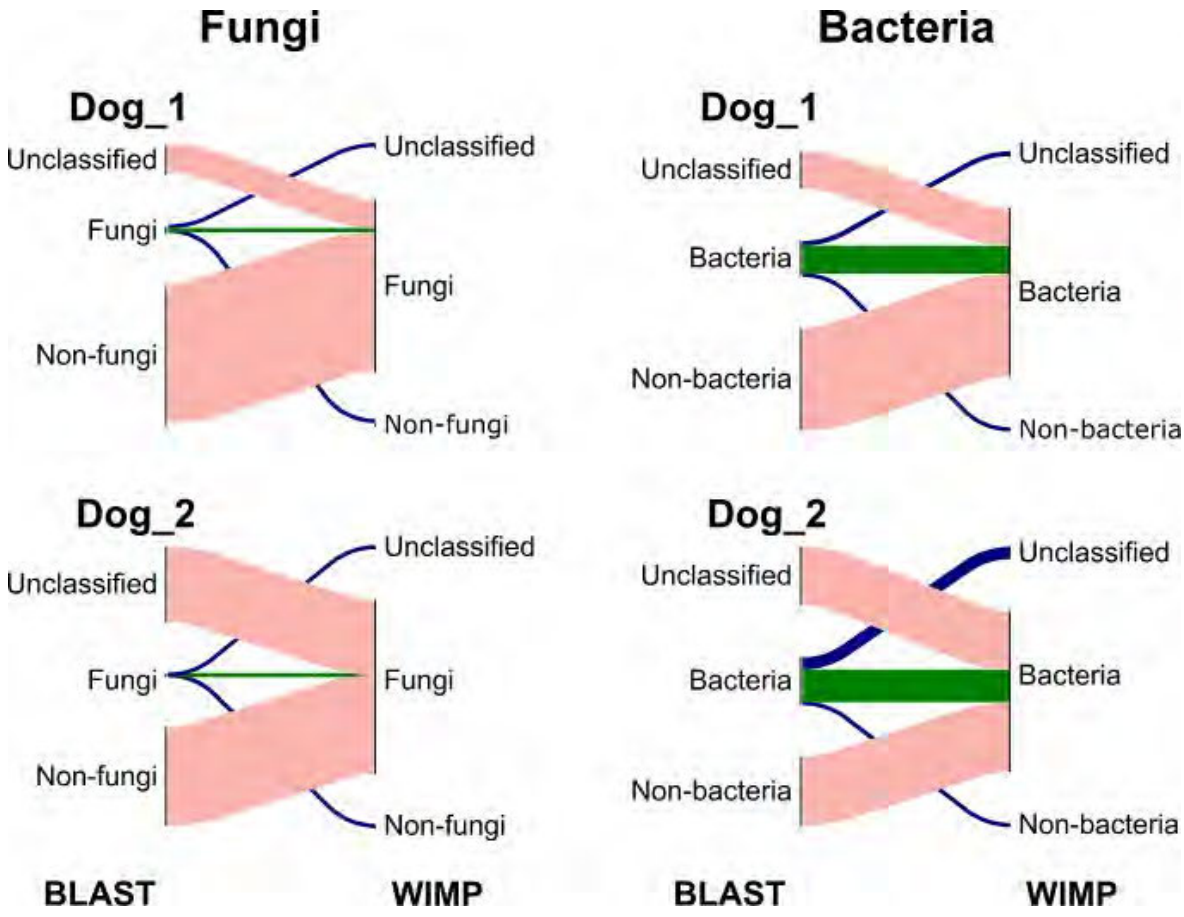


Figure 7.3. Comparison of fungal reads and bacterial reads classified by WIMP and BLAST from two dog BAL fluid samples. Green lines indicate species detection concordance, showing a much higher precision for bacteria than fungi, pink lines indicate WIMP and blue lines BLAST specific identifications.

7.6 Discussion

Pneumocystis is known to cause serious disease in human immunocompromised hosts but remains to be non-pathogenic within healthy individuals. Within dogs, the fungus unproportionally affects two breeds, Cavalier King Charles Spaniels (CKCS) and Miniature Dachshunds, even when they appear to be young, generally well and with no health issues [34, 38]. Due to this, a quick yet accurate diagnosis and commencing treatment as soon as

possible are essential for the survival of the dogs. The devastating nature of a PCP upon healthy dogs is a true concern to veterinarians and dog owners alike, and except for the work of the Danesi *et al.*, there have been minimal improvements for testing [31]. The importance of having rapid and accurate diagnosis is growing, and Nanopore MinION may offer a solution in the near future for rapid, multiplex testing solution directly from canine respiratory specimens. The current study provided a proof of concept that NGS using the Nanopore MinION can successfully identify *Pneumocystis* spp. in PCP positive dogs, from BAL specimens, when utilising the correct taxonomic reference algorithms, which in thiscase was BLAST.

BLAST is globally the most popular search algorithm used, which classifies read using an alignment-based system [39]. When observing the total readings obtained by the MinION, compared to the WIMP algorithm BLAST performed markedly better, and this remained throughout all analysis undertaken. This trend was also evident from the human PCP samples described in Irinyi *et al.* [40], Chapter 7.3. BLAST was unable to classify 25% of all reads for sample of Prince (Dog 1), and approximately 55% for the sample of Sophie (Dog 2),contrasted to WIMP which was unable to classify approximately 45% and 78% of Prince and Sophie, respectively. Surprisingly both algorithms found human reads amongst the totalnumber of reads identified in both dog samples. Using the BLAST algorithm, reads were distinguished as either microbial, human, dog, and other mammalian DNA, but when applying the WIMP algorithm, only human and microbial reads were able to be classified. Using the WIMP algorithm, significantly more microbial DNA was classified compared to BLAST. When comparing the concordance results of Figure 7.3, it is evident that the majority of DNA WIMP detected was bacterial and had a poor fungal classification. WIMP also detected a much larger quantity of human reads within the samples. The detection of human reads in both dog samples is probably due to the misclassification of dog DNA as

human DNA, opposed to the possibility of true human DNA comprising the majority of reads from a dog clinical sample. This could be due to the lack of dog specific bioinformatics algorithms. WIMP performed better when classifying reads from human PCP samples which is probably due to the large amount of annotated human genomes in the database. Well annotated dog genomes in the database should increase the accuracy of WIMP.

Observing only the classified fungal reads, using BLAST, the MinION was able to accurately classify *Pneumocystis,* to the genus level, as the major fungal pathogen within both samples. The samples of Prince (dog 1) had only 10% of the fungal reads identified not as *Pneumocystis*, and no other identifiable fungi, while Sophie's (dog 2) sample identified a total of 19 genera of fungi within her sample. From Sophie's sample, the second most abundant fungi classified were *Candida* spp., which may indicate a co-infection, or also just being commensal organisms from the respiratory or oral flora of the dog - it is unclear as it was not classified to the species level. Other genera of fungi identified in smaller amounts contain common fungi found within general environments of humans or other animals, such as *Saccharomyces, Aspergillus, Cryptococcus, Yarrowia, Schizosaccharomyces* or soil and plant fungi such as *Ascoidea, Fusarium, Leptosphaeria, Sporisorium, Tetrapisispora, Sporisorium, Verticillium* and *Wickerhamomyces.* These fungi may be a true representation ofcommensals of the dog, or contamination from everyday fungi when isolating the sample, which may have come from the dog, clinical space, or human error. Interestingly the fungi *Kwoniella* was also identified, which is a monotypic fungus, only found in the Bahamas and Florida wetlands. This is most likely indicative of a false positive opposed to being acontamination, which was also noted within human specimens. The WIMP algorithm detected a small amount of *Pneumocystis* only within Prince's sample. Larger amounts of plant fungi were instead detected, as well as *Lobosporangium* and *Paracoccidioides*. The latter two are interesting as both fungi are endemic only to the Americas, so the possibility of

the true fragments of these fungi are highly unlikely and are almost certainly false positives. A similar result was also seen within the human samples, which was subsequently determined to be a false positive [40]. A componence of the detection of fungi between the two algorithms was undertaken, and just as documented with human respiratory samples, the level of concordance of the obtained classification within fungi was low, but high within the bacterial community. It is theorised that the false positives, low fungal classification, and low concordance are most likely due to poor fungal reference genomes and lack of those in the databases, general likeness to other eukaryotic, as well as the high recombination of genes within fungi which may lead to cross-species contamination [41-43]. This remains a challenge for the use of metagenomic based NGS for taxonomic profiling of fungi, especially if identification at the species level is necessary. Another factor which may have impacted classification and identification of *Pneumocytis canis* is that at the time of analysis, the entire genome had not yet been sequenced, so the reference library would have been from the genome of *P. jirovecii, P. carinii and P. murina* [23, 25]. As *Pneumocystis* species have a strict host specificity, as such the lack of a reference genome may have significantly impactedthe ability of sequences to be accurately detected, as the various species of *Pneumocystis* can carry large genetic differences of up to 15%, not often seen in fungi [23, 44].

Independently to the algorithm used, Prince's samples classified more total reads, and specific *Pneumocystis* reads than Sophie's sample. The differences may be attributed to the quality of the respiratory samples. The samples of Prince were noted by the pathologist to be higher in volume than that of Sophie, and were not contaminated with blood, as Sophies were. Additionally, Prince reportedly had much more severe disease, and was five years older than Sophie. It is also possible that when the respiratory sample was collected, there was a higher environmental contamination from the veterinarian's surgery in Sophies sample. This highlights the importance of obtaining high quality respiratory samples, in sterile conditions

when possible, so that the accuracy of fungal classifications is improved. Since the samples were sequenced retrospectively, years after the infection, it is unclear at what stage of their illness the samples were collected, and whether treatment had already begun. If TMP-SMX had already been administered, it may result in limiting the amount of reads able to be detected by MinION sequencing.

When comparing the two canine samples, and three human samples from Chapter 7.3, the canine samples obtained a more confident confirmation of *Pneumocystis* and had this been done in a real-life clinical setting, a rapid diagnosis from the metagenomics NGS could confidently diagnose PCP and allow the clinicians to administer the correct treatment in a timelier manner. Future studies should aim to collect greater amounts of respiratory specimens from suspected and confirmed PCP infections to be sequenced by shotgun metagenomics NGS for the confirming the accuracy obtained, as this study was limited to only two canine samples.

The major limitations of the MinION sequencing identified first by Irinyi *et al.*, [40] and again evident within the canine study, they combine a combination of the specimen quality, low microbial DNA to host DNA ratio, poor fungal reference databases, and a lack ofspecific bioinformatic algorithms used. Improvements to the reference databases can be madepossible by increasing the submissions of fungal DNA sequences, but especially additions of *Pneumocystis* DNA from various hosts. Considering the potential, addressing these pitfalls could see shotgun metagenomics based NGS, on portable sequencing platforms, such as the MinION, to become a routine in both human and animal diagnostics.

7.7 References

1. Kchouk, M., J.-F. Gibrat, and M. Elloumi, *Generations of sequencing technologies: from first to next generation.* Biology and Medicine, 2017. **9**(3).

2. Reuter, J.A., D.V. Spacek, and M.P. Snyder, *High-throughput sequencing technologies.* Molecular cell, 2015. **58**(4): p. 586-597.

3. Gu, W., S. Miller, and C.Y. Chiu, *Clinical metagenomic next-generation sequencing for pathogen detection.* Annual Review of Pathology: Mechanisms of Disease, 2019. **14**: p. 319-338.

4. Langelier, C., *et al.*, *Metagenomic sequencing detects respiratory pathogens in hematopoietic cellular transplant patients.* American journal of respiratory and critical care medicine, 2018. **197**(4): p. 524-528.

5. Greninger, A.L., *et al.*, *Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis.* Genome medicine, 2015.**7**(1): p. 1-13.

6. Long, Y., *et al.*, *Diagnosis of sepsis with cell-free DNA by next-generation sequencing technology in ICU patients.* Archives of medical research, 2016. **47**(5): p. 365-371.

7. Han, D., *et al.*, *mNGS in clinical microbiology laboratories: on the road to maturity.* Critical reviews in microbiology, 2019. **45**(5-6): p. 668-685.

8. Charpentier, E., *et al.*, *Added value of next-generation sequencing for multilocus sequence typing analysis of a Pneumocystis jirovecii pneumonia outbreak.* Emerging infectious diseases, 2017. **23**(8): p. 1237.

9. Cummings, L.A., *et al.*, *Clinical next generation sequencing outperforms standard microbiological culture for characterizing polymicrobial samples.* Clinical chemistry, 2016. **62**(11): p. 1465-1473.

10. Barrett, S.R., *et al.*, *Sensitive Identification of Bacterial DNA in Clinical Specimens by Broad-Range 16S rRNA Gene Enrichment.* Journal of Clinical Microbiology, 2020. **58**(12): p. e01605-20.

11. Miao, Q., *et al.*, *Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice.* Clinical Infectious Diseases, 2018. **67**(suppl_2): p. S231-S240.

12. Gardy, J.L. and N.J. Loman, *Towards a genomics-informed, real-time, global pathogen surveillance system.* Nature Reviews Genetics, 2018. **19**(1): p. 9-20.

13. Wang, C., *et al.*, *Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance.* Genome research, 2007. **17**(8): p. 1195-1201.

14. Salipante, S.J., *et al., Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology.* Journal of clinical microbiology, 2015. **53**(4): p. 1072-1079.

15. Judge, K., *et al., Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes.* Journal of Antimicrobial Chemotherapy, 2015. **70**(10): p. 2775-2778.

16. Salipante, S.J., *et al., Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling.* Applied and environmental microbiology, 2014. **80**(24): p. 7583-7591.

17. Quail, M.A., *et al., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC genomics, 2012. **13**(1): p. 1-13.

18. Zhang, C., *et al., Value of mNGS in sonication fluid for the diagnosis of periprosthetic joint infection.* Arthroplasty, 2019. **1**(1): p. 1-10.

19. Jain, M., *et al., The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community.* Genome biology, 2016. **17**(1): p. 1-11.

20. Chiu, C., *et al., Rapid metagenomic identification of viral pathogens in clinicalsamples by real-time nanopore sequencing analysis.* 2015.

21. Judge, K., *et al., Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology.* Microbial Genomics, 2016. **2**(9).

22. Alanio, A., *et al., Diversity of Pneumocystis jirovecii during Infection Revealed by Ultra-Deep Pyrosequencing.* Frontiers in Microbiology, 2016. **7**(733).

23. Cissé, O.H., *et al., Comparative Population Genomics Analysis of the Mammalian Fungal Pathogen Pneumocystis.* mBio, 2018. **9**(3): p. e00381-18.

24. Fakruddin, M., *et al., Pyrosequencing: A next generation sequencing technology.* World Appl Sci J, 2013. **24**(12): p. 1558-1571.

25. Cissé, O.H., M. Pagni, and P.M. Hauser, *De novo assembly of the Pneumocystis jirovecii genome from a single bronchoalveolar lavage fluid specimen from a patient.* MBio, 2012. **4**(1): p. e00428-12.

26. Camargo, J.F., *et al., Next Generation Sequencing of Microbial Cell-Free DNA for Rapid Noninvasive Diagnosis of Infectious Diseases in Immunocompromised Hosts.* Biology of Blood and Marrow Transplantation, 2019. **25**(3, Supplement): p. S356-S357.

27. Zhang, Y., *et al., A cluster of cases of pneumocystis pneumonia identified by shotgun metagenomics approach.* Journal of Infection, 2019. **78**(2): p. 158-169.

28. Ma, L., *et al.*, *Genome analysis of three Pneumocystis species reveals adaptation mechanisms to life exclusively in mammalian hosts.* Nature communications, 2016. **7**(1): p. 1-14.

29. Farrow, B., *et al.*, *Pneumocystis pneumonia in the dog.* Journal of comparative pathology, 1972. **82**(4): p. 447-453.

30. Kanemoto, H., *et al.*, *Common variable immune deficiency in a Pomeranian with Pneumocystis carinii pneumonia.* Journal of Veterinary Medical Science, 2015: p. 14-0520.

31. Danesi, P., *et al.*, *Molecular diagnosis of Pneumocystis pneumonia in dogs.* Medical mycology, 2017. **55**(8): p. 828-842.

32. Watson, P., *et al.*, *Immunoglobulin deficiency in Cavalier King Charles spaniels with Pneumocystis pneumonia*. Journal of veterinary internal medicine, 2006. **20**(3): p. 523-527.

33. Sykes, J.E., *Canine and feline infectious diseases-E-BOOK*. 2013: Elsevier Health Sciences.

34. Ralph, E., *et al.*, *Pneumocystis canis pneumonia in dogs.* Microbiology Australia, 2015. **36**(2): p. 79-82.

35. Lobetti, R., A. Leisewitz, and J. Spencer, *Pneumocystis carinii in the miniature dachshund: case report and literature review*. Journal of small animal practice, 1996. **37**(6): p. 280-285.

36. Vieson, M.D., P. Piñeyro, and T. LeRoith, *A review of the pathology and treatment of canine respiratory infections.* Veterinary Medicine: Research and Reports, 2012. **3**: p. 25.

37. Canfield, P., D. Church, and R. Malik, *Pneumocystis pneumonia in a dog.* Australian Veterinary Practitioner, 1993. **23**(3): p. 150-154.

38. Kirberger, R.M. and R.G. Lobetti, *Radiographic aspects of Pneumocystis carinii pneumonia in the miniature Dachshund.* Veterinary Radiology & Ultrasound, 1998. **39**(4): p. 313-317.

39. Boratyn, G.M., *et al.*, *BLAST: a more efficient report with usability improvements.* Nucleic acids research, 2013. **41**(W1): p. W29-W33.

40. Irinyi, L., *et al.*, *Long-read sequencing based clinical metagenomics for the detection and confirmation of Pneumocystis jirovecii directly from clinical specimens: A paradigm shift in mycological diagnostics.* Medical mycology, 2020. **58**(5): p. 650-660.

41. Lindgreen, S., K.L. Adair, and P.P. Gardner, *An evaluation of the accuracy and speed of metagenome analysis tools.* Scientific reports, 2016. **6**(1): p. 1-14.

42. Sczyrba, A., *et al.*, *Critical assessment of metagenome interpretation—a benchmark of metagenomics software.* Nature methods, 2017. **14**(11): p. 1063-1071.

43. Merchant, S., D.E. Wood, and S.L. Salzberg, *Unexpected cross-species contamination in genome sequencing projects.* PeerJ, 2014. **2**: p. e675.

44. Edman, J.C., *et al.*, *Ribosomal RNA sequence shows Pneumocystis carinii to be a member of the fungi.* Nature, 1988. **334**(6182): p. 519-522.

Chapter 8

Discussion

8.1 Introduction

The *Pneumocystis* started to become a public health concern in the early 1980's when seemingly healthy individuals were coming in with a type of pneumonia infection called Pneumocystis Pneumonia (PCP), previously only rarely seen in severely sick patients [1]. In the next two decades following, it became the most common AIDS defying illness, although in recent times it is also a major health concern for organ transplant patients, patients with autoimmune diseases and malignancies [2]. Outbreaks and nosocomial infections of *Pneumocystis jirovecii* in the transplant setting have resulted in wide-scale graft loss and unnecessary death, yet despite this, the fungus remains poorly studied, in comparison to other pathogenic fungi causing infectious in humans and animals [3]. *Pneumocystis* is unable to grow and be propagated in pure culture and has a strict host specificity preventing the use of animal models, both fundamental impacting factors which have hindered research [4, 5]. The development of molecular methods has allowed researchers to work through this hurdle to gain a better inside into its transmission, epidemiology, and pathogenicity, but many aspects of this fungus remain to be investigated. A major pitfall facing *Pneumocystis* researchers is that there is little concordance and global standardization between research groups in how molecular methods are undertaken. This is especially evident in genotyping and epidemiological investigations of *Pneumocystis* spp., frequently inhibiting researchers tocompare data. This thesis aimed to identify and address the holes in the current methodology used for genotyping of *Pneumocystis* spp., in particular *P. jirovecii*, with hopes that improvements discussed within this work can bring the *Pneumocystis* scientific community closer towards global collaborations to propel epidemiological investigations and uncover solution to many of the questions behind this organism.

8.2 Optimization of DNA extraction

To undertake this task, optimization was first targeted from the very start of clinical investigations, which is the DNA extraction. To date, there had been no comparison published of which kits and extraction methods were most effective when obtaining fungal DNA directly from lower respiratory samples. A search was conducted of commonly used or well performing commercial kits in publications, and five were selected to be evaluated. The kits tested were the Norgen Sputum DNA Isolation kit, Qiagen QIAamp DNA mini kit, Qiagen QIAamp DNA blood kit, Zymo Research Quick-DNA Fungal/Bacterial miniprep kit, and the Molzym Ultra Deep Microbiome PrepEnrichment kit. All five are extraction kits, except the Molzym kit, are also enrichment kits which removed host DNA. All 5 kits were tested on a range of 9 *P. jirovecii* and 3 *P. canis* positive samples under three conditions. Except for the Molzym kit, which generated little to no DNA, it was found that the additionof a mechanical liquid nitrogen grinding step prior to commencing the kit as per manufacturer's instructions generated significantly more DNA. Testing the kits with the addition of saponin, which was found to be highly effective in bacterial samples to reduce host DNA, led to the total loss of *P. jirovecii* DNA. It is possible that the removal of host DNA also removed eukaryotic fungal cells, minimising DNA to below detectable levels.After the comparison of DNA concentration and PCR amplification, the QIAamp DNA Mini Kit with the addition of a liquid nitrogen grinding step delivered the strongest results. As a result, io this studies it is advised that moving forward, a versatile chemical lysis kit with the addition of a mechanical lysis step should be used when extracting *Pneumocystis* DNA from lower respiratory samples.

8.3 Optimization of the original ISHAM MLST scheme

Following the optimisation of the DNA extraction conditions, the optimisation of the MLST schemes was undertaken, as it is currently the accepted gold standard for genotyping *P. jirovecii* [6]. The scheme used officially by the International Society for Human and Animal Mycology (ISHAM) and promoted via the ISHAM MLST database, available at https://mlst.mycologylab.org/. The scheme used four genetic loci *β-TUB, DHPS,* ITS and *mt26S,* carried a high discriminatory power, and successfully traced an outbreak of PCP within the renal ward amongst organ transplant patients in NSW, Australia [3]. Despite this, the scheme had not been used frequently in other studies, as its PCR protocols were lengthy, complex, different for all four loci and the used primers were unable to amplify strong DNA bands form all studied samples. As such the scheme was herein revisited and the optimised new MLST scheme was than applied to PCP positive clinical samples. PCR protocols were initially tweaked, but there were little improvements in the amplification, so new primers were designed, including five new primer combinations for *β-TUB,* and two novel primer pairs for *mt26S*, *DHPS* and ITS. The primer sets were tested at varying annealing temperatures to detect the most optimal conditions. Two notable large differences betweenthe amplification protocols were that ITS was no longer amplified using nested-PCR, but utilised traditional PCR as the remaining loci, and the double amplification cycle was removed for the *DHPS* locus. These two aspects of the previous primers were time consuming and the new primer pairs were designed to perform without them. Comparisons ofthe achieved amplification and sequencing rates revealed the best possible primers, accompanied by a standardised PCR protocol for all four loci.

The primer pairs recommended for the amplification and sequencing of the four genetic loci, *β-TUB, DHPS,* ITS, *mt26S,* of the improved ISHAM MLST scheme for *P. jirovecii* were:

β-TUB	Forward primer: BTubulinF 5´-TCATTAGGTGGTGGAACGGG-3´ Reverse primer: PneumoTubReverseR3 5´- AAACGGCACCATATTAACGG -3´
DHPS	Forward primer: PneumoDHPSF1 5´- AGCAGTGCCCCAAATCCTAT -3´ Reverse primer: PneumoDHPSR2 5´-GCGCCTACACATATTATGGCCATTTTAAATC -3´
ITS	Forward primer: PneumoITSF 5´- CCATTGCTGGAAAGTTGATCA -3´ Reverse primer: PneumoITSR 5´- TCGCCGTTACTAAGGGAATC -3´
mt26S	Forward primer: PneumoLSUForward1 5´- TCAGGTCGAACTGGTGTACG -3´ Reverse primer: PneumoLSUReverse1 5´- TGTTTCCAAGCCCACTTCTT -3´

The protocol for the DNA extraction and the protocol for the PCR amplifications were previously very different, which meant that samples could not be prepared or amplified concurrently, making it overly time consuming, and increasing the likelihood of error and contamination. The working protocol for the PCR amplification was subsequently unified for all four loci, so all samples could be processed and amplified concurrently, resulting in a faster analysis. To trial the scheme on a real cohort, this new protocol was applied to 50 Chilean P. jirovecii positive clinical samples for genotyping. The scheme managed to successfully sequence all four loci from 23 of samples and uncovered one new *mt26S* allele type (AT6) and four new ITS allele types (AT8, AT9, AT10 and AT11). The 23 samples resulting in 16 sequence types, displaying a high discriminatory power of the scheme. Unfortunately, the most versatile locus, ITS, was unable to be sequenced in 50% of allstudied clinical samples, despite the optimisation undertaken, confirming finding reported in other publications [6, 7].

8.4 Establishment of a new ISHAM Consensus MLST scheme

The low sequencing success rates of this opimised MLST scheme deemed it inappropriate to suggest it as a consensus MLST scheme to the scientific community, therefore, a redesign of the MLST was undertaken, starting with a comparison of all past MLST schemes and loci used for *P. jirovecii* genotyping. Within the review, 67 publications were identified, with a total of 19 different loci used, composing 32 different MLST schemes used globally from 1994 to July 2020. This huge number of different MLST schemes being used in the community demonstrated a clear lack of accordance between research groups. As a result of this analysis the most popular loci were identified to be *β-TUB, CYB, DHPS,* ITS1 and ITS2, *mt26S* and *SOD*, these loci were then tested on a larger cohort of a totalling 130 samples, to strengthen the reliability of amplification and increase the discriminatory power. The scheme which proved to have the highest H-index (a measure of discriminatory power) and successfully amplified all samples consisted of *β-TUB, CYB, mt26S* and *SOD.* The ITS locus again sequenced the least of the studied samples and was as such excluded for a subsequent consensus scheme. In addition, the *DHPS* locus was also not considered suitable, due to its low discriminatory power. The primers used for *β-TUB* and *mt26S* were the same as optimised for the original ISHAM scheme, while the primers used to sequence *SOD* and *CYB*were obtained from Maitte *et al.* 2013, due to their proven reliability in amplifying these two loci [6].

The new primers used for the amplification of the final MLST loci are:

β-TUB	Forward primer: BTubulinF 5´-TCATTAGGTGGTGGAACGGG-3´ Reverse primer: PneumoTubReverseR3 5´- AAACGGCACCATATTAACGG -3´
CYB	Forward primer: CytbFw 5'-CCCAGAATTCTCGTTTGGTCTATT-3' Reverse primer: CytbRw 5'-AAGAGGTCTAAAAGCAGAACCTCAA-3'
mt26S	Forward primer: PneumoLSUForward1 5´- TCAGGTCGAACTGGTGTACG -3´ Reverse primer: PneumoLSUReverse1 5´- TGTTTCCAAGCCCACTTCTT -3´
SOD	Forward primer: MnSODFw 5'-GGGTTTAATTAGTCTTTTTAGGCAC-3' Reserve primer: MnSODRw 5'-CATGTTCCCACGCATCCTAT-3'

The PCR protocols were also standardised, with the only difference between the loci are that the *β-TUB* and the *CYB* loci amplification protocols have a five degree increase in the annealing temperature.

8.5 Application of the new ISHAM Consensus MLST Scheme to a Cohort of *Pneumocystis jirovecii* positive clinical Samples

The newly developed consensus MLST scheme was then applied to a selected cohort of clinical samples obtained from a single hospital in Barcelona, Spain, that contained 86samples from both with PCP and *P. jirovecii* colonised individuals, with a variety of co- infections, and from different types of respiratory specimens. The samples were accompaniedby metadata, which made them a perfect sample set for the population genetic study of *P. jirovecii* in Barcelona between 2014 and 2018. The new consensus scheme was effectively applied this epidemiological study, discovering 32 sequence types, and even linking certain samples to cross transmission events within the hospital, which was not known prior. Using the scheme, researchers were also able to attribute the possible clinical relevance of certain

sequence types, such as ST3, ST30 and ST31, which were identified to appear exclusively in confirmed pneumonia patients, while ST27 only was associated with colonised individuals. The set of genetic loci, *β-TUB, CYB, mt26S* and *SOD*, was subsequently put forward as the consensus MLST scheme to be used in future genotyping studies of *P. jirovecii* to the global scientific and medical community, alongside the established online database at mlst.mycologylab.com [8, 9]. The utilisation of a consensus scheme and subsequent addition of the generated sequences into the centralised MLST database will allow global collaborations, large-scale population genetics studies, and the comparisons of the obtained results, to be undertaken as the basis for appropriate risk analysis an informed public heath response to nosocomial outbreaks.

A pilot study of 145 globally collected *P. jirovecii* positive clinical samples, originating from five countries across three continents was then undertaken to further prove the effectiveness of the new consensus MLST scheme. For the first time, samples from Brazil and New Zealand were sequenced. The global analysis showed that among the 145 sequences, 49 unique sequence types were identified, along with novel allele types (AT) identified among all four loci. Novel allele types discovered were AT3 and AT4 for the *β-TUB* locus, AT12 and AT13 from the *CYB* locus, AT9 and AT10 for the *mt26S* locus, well asAT4, AT5 and AT6 for the *SOD* locus. The analysis showed that each locus had dominantand minor allele types in circulation. Dominant alleles identified were AT1 and AT2 for the *β-TUB* locus, AT1 and AT4 of the *CYB* locus, AT2 and AT4 of the *mt26S* locus, and AT1 andAT2 for the *SOD* locus, which all where present across multiple countries and had been reported previously [6, 10, 11]. Unsurprisingly the dominant allele types also gave rise to the globally dominant sequence types detected, ST3, ST5, ST13, which were all present across different countries. Conversely, all novel genotypes identified were detected in minute numbers, and are geographically restricted to the country they have been identified from. It is

possible that globally distributed variants remain dominant within circulation for many years, but certain geographically specific pressures within countries or environments give rise to variants which circulate within smaller populations. Interestingly, the global analysis detected a shift in allele dominance in circulation reported in previous years, for example the *CYB* AT4 which was only previously detected in minute numbers, now made up 15% of all *CYB* sequences identified.

These findings raised more questions than were managed to be answered during this study. These include: (1) Are allele types recent *de novo* mutations or have they beencirculating within the communities for much longer but were never identified before? (2) What role do colonised individuals play in the spread; and (3) Are some genetic variantsmore clinically relevant than others? A number of limitations of the current study, such as small sample size from some countries, retrospective clinical samples opposed to samples from recent (<1 year) infections, and the lack of metadata, do not yet allow for thesequestions to be addressed. But the pilot study showed that large scale population screening allows for fragments of the *Pneumocystis* puzzle to be uncovered.

The creation of a centralized database dedicated to only one scheme allows to for the controlled addition of *P. jirovecii* MLST loci sequences from clinicians and researchers globally. If sequences are added promptly into the database, this will allow all *Pneumocystis* researchers access to an up-to-date current genotype sequence date set from clinical samples, opposed to using retrospective samples, waiting for publications to be published and independently searching and collating literature to access this data. Hopefully with the uptake of the herein developed and proposed consensus MLST scheme more additions will be made into the database, making a wide scale global screening on a mass level never undertaken before possible.

8.6 *DHPS* as a genetic locus to monitor mutation conferring resistance

It is important to note that *DHPS*, a valuable locus, was removed from the consensus scheme, despite appearing in the bulk of previous *Pneumocystis* typing publications. *DHPS* was not deemed to be appropriate for distinguishing between various strains of *P. jirovecii*, but that is not to say that researchers should stop sequencing this locus. *DHPS* has been used to monitor potential resistance towards primary treatment of PCP, TMP-SMX, since the 1990's, when it was noticed that the high use of sulfa drugs led due to the increase of mutations within codons 55 and 57 [12, 13]. It is unclear which clinical relevance this has on patients, with many conflicting reports in the literature suggesting mutations may cause treatment failure, and increase risk of serious disease, ICU and ventilation need, and even an increased risk of death, while other reports show no significant correlation [14-19]. Additionally, populations with no previous sulfa exposure were detected harbouring mutations, suggesting other environmental or geographical pressures influencing significant mutations [20-22]. There have been no large-scale population studies which were able to confidently draw conclusions about the impact of mutations and what drives them. In the past decades, there has been a decline of mutations detected, largely attributed to HAART therapy, but mutations are still found within population with no clear sulfa link, and two recent studies detected mutations within other codons [10, 23, 24]. This outlines the importance of continuing to monitor changes within *DHPS*, and once again, a centralised database for *DHPS* allele submissions is vital, so wide-scale monitoring of mutations can be up-to-date, collected, and collated. The ISHAM database also accepts submissions for the *DHPS* locus, separate to that of the MLST scheme, for the very reason to feat a large-scale population analysis to better understand the driving forces and clinical implications of mutations.

In-house analysis of the *DHPS* loci submitted within the ISHAM database from Australia, Brazil, Chile, and New Zealand was conducted, and found that mutations are only evident in Chile and Brazil. Previously a report from Chile detected high levels of mutations within the community, including in patients with no prior sulfa exposure, and this recent analysis showed that 10% of samples were mutants at codons 55 and 57 [21]. The mutation rates in Chile were unproportionally higher than the one detected in neighbouring countries, while TMP-SMX use remained similar, suggesting population pressure from high sulfa use was not the only impacting factor. As the sample set from Chile was collected from infections over five years after the last publication, the results may indicate a downward trend of mutations. Conversely a mutation in the Brazilian sample set was identified, which had not been previously reported in Brazilian populations. Unfortunately, due to the small sample size and the lack of metadata available, no clinical relevance could be concluded, nor could any trends be confidently identified. This will continue to be a hurdle researchers face until an avenue for the collation of data is utilised.

Another limitation identified in current studies is, that most publications only survey samples for mutations in codons 55 and 57, while other mutations are rarely investigated, especially if they are synonymous or silent mutations. In all other loci genotyped from *P. jirovecii* all variants with no 55 and 57 mutations are regarded as wildtype and grouped together, despite if they are harbouring any other synonymous mutations or not, which is not a nomenclature adapted by any other loci genotyped. In our investigation, by noting changes outside of the two domains, three novel variants of the wild type/allele type 1 were identified. Interestingly, the sample from New Zealand from which the AT1 v3 was detected, also harboured a rare mutation in the *mt26S* locus, while the same sample with the *DHPS* AT1 v2 from a Brazil also harboured a novel mutation in the *CYB, mt26S,* and *SOD* loci. Since the results of two samples are not enough to draw any strong conclusions, it is advisable that

researchers continue to be independently sequence *DHPS* whenever possible and that all variations within the locus are detected and allocated. Exploring all *DHPS* mutations may give insights into the underlying causes or impacts of mutations not previously known. WhenMLST genotyping is not possible, clinicians may want to consider sequencing the *DHPS* locus for identification purposes. All samples amplified within the herein presented workwere successfully sequenced after the primers and PCR protocols had been optimised. This will allow for both the detection and quick monitoring of any mutations harboured to monitorany potential emerging resistance.

8.7 NGS as diagnostic tool for the diagnosis of *Pneumocystis* infections in humans and dogs

The final leg of the herein reported studies exploring sequencing of *Pneumocystis* steps away from traditional way of generating sequences, as the future directions of scientific diagnostic and epidemiological studies appear to be headed towards next-generation sequencing. A handful of *Pneumocystis* studies have utilised third generation sequencing technology, such as pyrosequencing, but little-to-no work had been undertaken exploring more modern techniques, such as shotgun metagenomic next generation sequencing [5, 25]. The Nanopore MinION had been dubbed a promising platform for the diagnosis of infectious diseases, as it is able to obtain results directly from clinical samples and could be connected to any portable computer to become a new tool to be included in bench to bedside diagnosis. To contact a pilot study five positive *Pneumocystis* samples, three from humans and two from dogs, were subjected to metabarcoding based NGS studies using the MinION instrument and were successfully identified as either *P. jirovecii* or *P. canis*. This is a highly exciting finding, as both species rely on a timely diagnosis to improve survival rates, besides the organisms being notoriously difficult to diagnose via microscopic detection [26]. A major limitation at the

time of the study was that there was no known *P. canis* genome to use as a reference, but despite this, the dog samples run through the MinION, using the BLAST algorithm, provided more classified *Pneumocystis* reads than the human samples.

Since the proof of concept was published in late 2019, additional five case reports and studies have been published demonstrating the use of metagenomics based NGS for the identification of PCP, as well as the ability to detect unknown co-infections, such as with cytomegalovirus or *Aspergillus fumigatus* contributing to their declining health, from respiratory samples [27-30]. In one case, metagenomics based NGS was able to accurately determine *Pneumocystis* as the causative agent of pneumonia in a renal transplant unit, despite not being diagnosed with PCP due to unsuccessful microscopy. When all exhaustive tests were negative, a BAL sample was sent to Beijing Genomics Institute and within 72 hours it had been discovered that *P. jirovecii* accounted for 99% of all microbial reads [29]. Although all studies applied metagenomics based NGS, none explored the use of the MinION for rapid bedside testing, but all studies sent clinical specimens to commercial sequencing companies. To truly change the face or rapid and accurate diagnostics, a cost and time effective platform needs to be utilised, so that outsourcing is not needed, and all reads can be classified from within the hospital or adjunct microbiology diagnostic laboratory. The advances made in the past two years show serious promises of rapid bedside testing one day becoming a reality. Unfortunately, neither our results, nor the one from the other publications, were able to incorporate genotyping whilst diagnosing PCP using metagenomics based NGS, as no targeted gene regions could be successfully identified tothat degree.

The value of metagenomics based NGS for the diagnosis of PCP should continue to be assessed in future studies. As more *Pneumocystis* sequences accumulate in reference

databases, the speed and accuracy of metagenomics based NGS will subsequently improve. The studies to date have focused on the positive diagnosis from a handful of cases, but no large-scale screening has been undertaken. Large scale screening, especially with the inclusions of both PCP and colonised samples, which could give rise to quantitative thresholdvalues able to distinguish between the two and widescale population analysis to understand the role colonised humans (and dogs) may play in the transmission and source for PCP infections. Rapid bedside testing of PCP patients could also allow for other studies, such as the monitoring of prophylaxis or treatments regimes, and the relevance in clinical outcomes. For canine PCP infections, metagenomics based NGS can now be more accurately undertaken than it was for our pilot study, with the successful sequencing of the *P. canis* genome in mid-2021, enriching the reference databases for increased sensitivity and specificity [31].

8.8 General conclusions

The underlying theme of the herein presented work is, that it highlights the desperate need for greater concordance within the *Pneumocystis* community, a scientific community which, even in 2021, are in discordance about the nomenclature being used for Pneumocystis Pneumonia and with no official standardisation of research protocols between laboratory groups [32]. Hundreds of publications discussing the epidemiology of this fungus have been produced, but without allowing for data sharing, investigations moving forward will continue to be limited and stunted. The studies presented herein have considered past research undertaken, techniques and procedures, reviewed them, and tried, tested and put forward optimised protocols for DNA extractions, PCR primers and protocols, as well as an optimised, global consensus MLST scheme on association with a centralised MLST sequencedatabase. This should allow for research groups to compare their obtained results,

collaborate, and allow answer many still open questions, which can only be attempted with wide-scale population analysis. This study encourages all *Pneumocystis* researchers to adopt a level of uniformity within their practices and collaboratively propel *P. jirovecii* epidemiology to levels of other well-known infectious fungi and bacteria.

8.9 References

1. Mocroft, A., *et al.*, *Changes in AIDS-defining illnesses in a London Clinic, 1987- 1998.* Journal of acquired immune deficiency syndromes (1999), 1999. **21**(5): p. 401- 407.

2. Kelly, M.N. and J.E. Shellito, *Current understanding of Pneumocystis immunology.* Future microbiology, 2010. **5**(1): p. 43-65.

3. Phipps, L.M., *et al.*, *Nosocomial Pneumocystis jirovecii pneumonia: lessons from a cluster in kidney transplant recipients.* Transplantation, 2011. **92**(12): p. 1327-1334.

4. Liu, Y., G.A. Fahle, and J.A. Kovacs, *Inability to culture Pneumocystis jirovecii.* MBio, 2018. **9**(3): p. e00939-18.

5. Cissé, O.H., *et al.*, *Comparative Population Genomics Analysis of the Mammalian Fungal Pathogen Pneumocystis*. mBio, 2018. **9**(3): p. e00381-18.

6. Maitte, C., *et al.*, *Multilocus sequence typing of Pneumocystis jirovecii from clinical samples: how many and which loci should be used?* Journal of clinical microbiology, 2013. **51**(9): p. 2843-2849.

7. Totet, A., *et al.*, *Pneumocystis jirovecii genotypes and granulomatous pneumocystosis.* Médecine et maladies infectieuses, 2006. **36**(4): p. 229-231.

8. Pasic, L., *et al.*, *Consensus Multilocus Sequence Typing Scheme for Pneumocystis jirovecii.* Journal of Fungi, 2020. **6**(4): p. 259.

9. L. Goterris, L.P., M. Guerrero Murillo, A. Kan, A. Anton, J. Aguilar Company, I. Ruiz-Camps, W. Meyer, M.T. Martin-Gomez, *Pneumocystis jirovecii geneticdiversity in a Spanish tertiary hospital.* . Medical mycology, 2021. **Accepted on the 29/8/2021**.

10. Esteves, F., *et al.*, *Pneumocystis jirovecii multilocus genotyping profiles in patients from Portugal and Spain.* Clinical microbiology and infection, 2008. **14**(4): p. 356- 362.

11. Desoubeaux, G., *et al.*, *Epidemiological Outbreaks of Pneumocystis jirovecii Pneumonia Are Not Limited to Kidney Transplant Recipients: Genotyping Confirms Common Source of Transmission in a Liver Transplantation Unit.* Journal of Clinical Microbiology, 2016. **54**(5): p. 1314-1320.

12. Kazanjian, O., *et al., Dihydropteroate synthase mutations in Pneumocystis carinii associated with sulfa prophylaxis failures.* AIDS, 1998. **12**: p. 873-78.

13. Kazanjian, P., *et al., Pneumocystis carinii mutations are associated with duration of sulfa or sulfone prophylaxis exposure in AIDS patients.* The Journal of infectious diseases, 2000. **182**(2): p. 551-557.

14. Takahashi, T., T. Kanda, and A. Iwamoto, *Genetic diversity of drug targets including dihydropteroate synthase, dihydrofolate reductase and cytochrome b, in Pneumocystis carinii f. sp. hominis isolates in Japan.* Res Commun Mol PatholPharmacol, 2002. **112**(1-4): p. 159-76.

15. Navin, T.R., *et al., Effect of mutations in Pneumocystis carinii dihydropteroate synthase gene on outcome of P carinii pneumonia in patients with HIV-1: a prospective study.* The Lancet, 2001. **358**(9281): p. 545-549.

16. van Hal, S.J., *et al., Clinical significance and phylogenetic relationship of novel Australian Pneumocystis jirovecii genotypes.* Journal of clinical microbiology, 2009. **47**(6): p. 1818-1823.

17. Alvarez-Martínez, M.J., *et al., Pneumocystis jirovecii pneumonia in Spanish HIV-infected patients in the combined antiretroviral therapy era: prevalence of dihydropteroate synthase mutations and prognostic factors of mortality.* Diagnostic Microbiology and Infectious Disease, 2008. **62**(1): p. 34-43.

18. Helweg-Larsen, J., *et al., Effects of mutations in Pneumocystis carinii dihydropteroate synthase gene on outcome of AIDS-associated P carinii pneumonia.* The Lancet, 1999. **354**(9187): p. 1347-1351.

19. Mei, Q. and S. Gurunathan, *Failure of co-trimoxazole in Pneumocystis carinii infection and mutations in dihydropteroate synthase gene.* The Lancet, 1998. **351**(9116): p. 1631-1632.

20. Latouche, S., *et al., Pneumocystisjirovecii dihydropteroate synthase genotypes in French patients with pneumocystosis: a 1998–2001 prospective study.* Medical Mycology, 2003. **41**(6): p. 533-537.

21. Ponce, C.A., *et al., High prevalence of Pneumocystis jirovecii dihydropteroate synthase gene mutations in patients with a first episode of Pneumocystis pneumonia in Santiago, Chile, and clinical response to trimethoprim-sulfamethoxazole therapy.* Antimicrobial agents and chemotherapy, 2017. **61**(2): p. e01290-16.

22. Huang, L., *et al., Sulfa or sulfone prophylaxis and geographic region predict mutations in the Pneumocystis carinii dihydropteroate synthase gene.* The Journal of infectious diseases, 2000. **182**(4): p. 1192-1198.

23. Suárez, I., *et al., Low prevalence of DHFR and DHPS mutations in Pneumocystis jirovecii strains obtained from a German cohort.* Infection, 2017. **45**(3): p. 341-347.

24. Singh, Y., *et al., Circulating genotypes of Pneumocystis jirovecii and its clinical correlation in patients from a single tertiary center in India.* European Journal of Clinical Microbiology & Infectious Diseases, 2017. **36**(9): p. 1635-1641.

25. Alanio, A., *et al.*, *European study on Pneumocystis jirovecii short tandem repeats genotyping reveals wide population diversity with geographic specificities.* Journal de Mycologie Médicale, 2016. **26**(2): p. e5-e6.

26. Danesi, P., *et al.*, *Molecular diagnosis of Pneumocystis pneumonia in dogs.* Medical mycology, 2017. **55**(8): p. 828-842.

27. Jiang, J., *et al.*, *Metagenomic Next-Generation Sequencing for the Diagnosis of Pneumocystis jirovecii Pneumonia in Non-HIV-Infected Patients: A Retrospective Study.* Infectious Diseases and Therapy, 2021: p. 1-13.

28. Li, Y., *et al.*, *Application of metagenomic next-generation sequencing for bronchoalveolar lavage diagnostics in critically ill patients.* European Journal of Clinical Microbiology & Infectious Diseases, 2020. **39**(2): p. 369-374.

29. Chen, J., *et al.*, *Metagenomic next-generation sequencing in diagnosis of a case of pneumocystis jirovecii pneumonia in a kidney transplant recipient and literature review.* Infection and Drug Resistance, 2020. **13**: p. 2829.

30. Wang, J., Y. Han, and J. Feng, *Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis.* BMC pulmonary medicine, 2019. **19**(1): p. 1-8.

31. Cissé, O.H., *et al.*, *Genomic insights into the host specific adaptation of the Pneumocystis genus.* Communications Biology, 2021. **4**(1): p. 305.

32. Nevez, G., *et al.*, *It is still PCP that can stand for Pneumocystis pneumonia: Appeal for generalized use of only one acronym.* Medical Mycology, 2021.

Appendix 1

Clinical Isolates Studied, Respective Metadata and Genotyping Results

Clinical Isolates Studied, Respective Metadata and Genotyping Results

	Metadata Available											Allele Type as per ISHAM consensus MLST Scheme (Pasic *et al.* 2020)					Allele type of Additional Loci (when available)	
Sample ID	**Other associated ID Codes**	**Date of Isolation**	**Date received**	**Specific source**	**Colonized or PCP**	**Supplier**	**Country**	**Comorbidities**	**Sex**	**PCP Positive**	**Age**	***β-TUB* AT**	***CYB* AT**	***mt26S* AT**	***SOD* AT**	**Sequence Type**	***DHPS***	**ITS**
62.1	1316457	8/11/2016	2017	IS		1	Australia	Renal Tx	M	Yes		1	1	4	2	3	1	N-S
62.2	1316457	8/11/2016	2017	IS		1	Australia	Renal Tx	M	Yes		1	1	4	1	13	1	N-A
63	392069	20/12/2106	2017	BAL & IS		1	Australia	Renal Tx	F	Yes		1	1	4	2	3	1	N-A
64.1	721593	1/11/2016	2017	IS		1	Australia	Renal Tx	M	Yes		1	1	4	1	13	1	N-S
64.2	721593	1/11/2016	2017	IS		1	Australia	Renal Tx	M	Yes		1	1	4	2	3	1	N-A
65	MB-16-408473	1/02/2017	2017	IS		2	Australia	Renal Tx	F	Yes		1	1	4	2	3	1	N-A
66	80-17-3-013-4717	12/12/2016	2017			2	Australia	lupus nephritis	F	Yes		2	4	2	2	5	1	N-S
67	MB-16-397491	20/12/2016	2017			2	Australia	Renal Tx	M	Yes		2	4	2	2	5	1	N-S
68	MB-17-3350	1/12/2016	2017			2	Australia	Renal Tx	F	Yes		1	1	4	1	13	1	N-S
	MB-17-4563	10/12/2016	2017			2	Australia	Renal Tx	F	Yes		N-A	N-P	2	N-P		1	N-P
	MB-18-411922		2018			2	Australia			Yes		N-S	N-P	4	N-P		1	N-P
	MB-18-416196		2018			2	Australia			Unknown		N-A	N-A	N-A	N-A		N-A	N-P
	MB-18-428819		2018			2	Australia			Yes		N-P	N-P	4	N-P		1	N-P
	MB-18-428819		2018			2	Australia			Yes		N-P	N-P	4	N-P		1	N-P
	MB-19180377		2019			2	Australia			Yes		N-P	N-P	N-P	N-P		1	N-P
73	73		2017			3	Brazil					N-A	N-A	N-A	N-A		N-A	N-A
76	76		2017			3	Brazil					3	12	8	5	14	1	N-P

85	85		2017			3	Brazil					1	13	10	6	11	1	N-P
87	87		2017			3	Brazil					2	2	4	5	2	1	N-P
88	88		2017			3	Brazil					1	13	10	6	11	1	N-P
89	89		2017			3	Brazil					N-S	N-S	1	4		1	N-P
103	103		2017			3	Brazil					1	4	1	1	18	1	N-P
104	104		2017			3	Brazil					3	12	8	5	14	1	N-P
105	105		2017			3	Brazil					1	8	3	2	19	1	N-P
107	107		2017			3	Brazil					1	4	1	1	18	1	N-P
108	108		2017			3	Brazil					1	8	3	2	19	1	N-P
109	109		2017			3	Brazil					4	12	5	6	10	3	N-P
111	111		2017			3	Brazil					1	8	4	1	15	1	N-P
112	112		2017			3	Brazil					1	8	3	2	19	1	N-P
113F	113F		2017			3	Brazil					N-S	13	9	N-S		1	N-P
115	115		2017			3	Brazil					1	4	1	1	18	1	N-P
119	119		2017			3	Brazil					1	13	10	6	11	1 v2	N-P
144	144		2017			3	Brazil					1	8	4	1	15	1	N-P
151	151		2018			3	Brazil					N-S	11	1	1		1	N-P
154	154		2017			3	Brazil					3	12	8	5	14	1	N-P
106F	106F		2017			3	Brazil					2	2	4	5	2	1	N-P
506	506		2016		PCP	4	Chile			Yes		N-A	N-P	2	N-P		1	N-A
509	509		2016		PCP	4	Chile			Yes		2	5	4	1	7	1	N-S
510	510		2016		PCP	4	Chile			Yes		1	1	4	2	3	1	N-A
515	515	15/12/2004	2016	BAL	PCP	4	Chile		M	Yes	34	1	8	5	1	4	1	7
524	524		2016		PCP	4	Chile			Yes	53	2	4	2	2	5	1	N-S
531	531		2016		PCP	4	Chile			Yes		1	5	6	6	12	1	N-S
1777	1777	17/01/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		1	9	5	1	6	1	8
1794	1794	17/02/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	23	1	1	4	2	3	1	8
1813	1813	31/03/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	55	1	1	4	2	3	4	8
1819	1819	6/04/2011	2016	BAL	PCP	4	Chile	Interstitial Pneumonia	F	Yes	55	1	9	5	1	6	1	8
1833.1	1833.1	6/05/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	55	2	5	4	1	7	3	9

1833.2	1833.2	6/05/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	60	2	8	4	1	8	3	9
1839	1839	18/05/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		2	5	2	6	9	1	8
1862	1862		2016		PCP	4	Chile			Yes		1	5	6	6	12	1	N-S
1887	1887		2016		PCP	4	Chile	HIV-positive	M	Yes	40	2	2	4	2	16	1	N-S
1889	1889	1/08/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	32	2	8	4	1	8	4	9
1915	1915		2016		PCP	4	Chile			Yes		1	1	4	2	3	1	N-S
1925	1925	5/12/2011	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	52	1	4	2	2	17	1	8
1955	1955		2016		PCP	4	Chile			Yes		2	4	2	2	5	1	N-S
1984	1984		2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		1	5	6	6	12	1	N-S
2028	2028	3/09/2012	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	33	1	1	4	2	3	1	9
2045	2045		2016		PCP	4	Chile			Yes		2	2	4	2	16	1	N-S
2060	2060		2016		PCP	4	Chile			Yes		2	2	4	2	16	1	N-S
2061	2061	15/01/2013	2016	BAL	PCP	4	Chile	HIV-positive		Yes	45	1	8	4	1	15	1	11
2081	2081		2016		PCP	4	Chile			Yes		1	1	4	1	3	1	N-S
2099	2099		2016		PCP	4	Chile			Yes		N-A	N-P	2	N-P		1	N-A
2103	2103	11/12/2013	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		1	8	4	1	15	1	10
2104	2104	11/12/2013	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	35	2	2	4	2	16	1	9
2107	2107	30/12/2013	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		1	8	4	1	15	1	10
2108	2108	7/01/2014	2016	BAL	PCP	4	Chile	HIV-positive	F	Yes		1	1	4	2	3	1	7
2110	2110		2016		PCP	4	Chile			Yes		N-S	N-P	4	N-P		1 v1	N-A
2113	2113		2016		PCP	4	Chile			Yes		2	4	2	2	5	1	N-S
2116.1	2116	6/05/2013	2016		PCP	4	Chile	HIV-positive	M	Yes		2	10	2	2	29	4	N-S
2116.2	2116	6/05/2013	2016		PCP	4	Chile	HIV-positive	M	Yes		2	4	2	2	5	1	N-S
2117	2117	19/02/2014	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	49	1	5	6	6	12	1	7
2124	2124	19/03/2014	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		2	5	2	6	9	1	8
2129.1	2129.1		2016		PCP	4	Chile			Yes		2	5	4	1	7	1	N-S
2129.2	2129.2		2016		PCP	4	Chile			Yes		2	8	4	1	8	1	N-S
2132	2132		2016		PCP	4	Chile			Yes		N-A	N-P	6	N-P		1	N-A
2135	2135		2016		PCP	4	Chile			Yes		N-A	N-P	2	N-P		1	N-S

2136	2136	16/05/2014	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	31	1	1	4	1	3	1	10
2142	2142		2016		PCP	4	Chile			Yes		N-A	N-P	5	N-P		1	N-S
2156	2156	30/10/2014	2016	BAL	PCP	4	Chile	HIV-positive	F	Yes	40	2	8	4	1	8	3	9
2159	2159		2016		PCP	4	Chile		M	Yes	52	1	10	2	2	1	1	N-S
2160	2160	25/11/2014	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	43	2	5	2	6	9	1	8
2162	2162	7/12/2014	2016	BAL	PCP	4	Chile	HIV-positive	F	Yes	46	1	1	4	2	3	1	8
2164.1	2164.1	12/01/2015	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	54	1	1	4	2	3	1	2
2164.2	2164.2	12/01/2015	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	54	2	4	2	2	5	1	2
2165	2165		2016		PCP	4	Chile			Yes		2	2	4	5	2	1	N-S
2166	2166	30/01/2015	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes		1	9	5	1	6	1	8
2171	2171	26/06/2015	2016	BAL	PCP	4	Chile	HIV-positive	M	Yes	63	1	8	5	1	4	1	N-S
HVH10.1	HVH10.1	5/08/2014		BAL	Colonisation	5	Spain	Hematopoyetic stem cell transpantation	M	No	64	2	1	5	1	37	1	
HVH10.2	HVH10.2	5/08/2014		BAL	Colonisation	5	Spain	Hematopoyetic stem cell transpantation	M	No	64	2	1	3	1	21	1	
HVH11.1	HVH11.1	23/09/2014		BAL	Colonisation	5	Spain	Lymphoma	F	No	43	1	6	5	3	45	1	
HVH11.2	HVH11.2	23/09/2014		BAL	Colonisation	5	Spain	Lymphoma	F	No	43	1	6	11	3	46	1	
HVH12	HVH12	24/10/2014		BAL	Colonisation	5	Spain	Hematopoyetic stem cell transpantation	M	No	78	1	8	4	1	15	1	
HVH13.1	HVH13.1	7/11/2014		BAL	Colonisation	5	Spain	Esophagic perforation	M	No	69	2	1	2	2	32	1	
HVH13.2	HVH13.2	7/11/2014		BAL	Colonisation	5	Spain	Esophagic perforation	M	No	69	2	1	3	2	22	1	
HVH14.1	HVH14.1	14/11/2014		BAL	Colonisation	5	Spain	Lung cancer	M	No	58	2	1	2	2	32	1	
HVH14.2	HVH14.2	14/11/2014		BAL	Colonisation	5	Spain	Lung cancer	M	No	58	2	1	3	2	22	1	
HVH16	HVH16	18/11/2014		BAL	Colonisation	5	Spain	ALL	F	No	59	2	1	2	2	32	1	

HVH17.1	HVH17.1	10/12/2014		BAL	Colonisation	5	Spain	AML	M	No	22	2	1	3	2	22	1	
HVH17.2	HVH17.2	10/12/2014		BAL	Colonisation	5	Spain	AML	M	No	22	2	1	3	3	47	1	
HVH18.1	HVH18.1	16/12/2014		BAL	Colonisation	5	Spain	Ricther syndrome post CLL	F	No	65	2	1	2	2	32	1	
HVH18.2	HVH18.2	16/12/2014		BAL	Colonisation	5	Spain	Ricther syndrome post CLL	F	No	65	2	1	3	2	22	1	
HVH19.1	HVH19.1	16/01/2015		BAL	Colonisation	5	Spain	AML	M	No	63	2	7	4	1	27	1	
HVH19.2	HVH19.2	16/01/2015		BAL	Colonisation	5	Spain	AML	F	No	63	2	7	4	3	50	1	
HVH21	HVH21	23/01/2015		BAL	Colonisation	5	Spain		F	No	49	2	7	4	1	27	1	
HVH22	HVH22	26/01/2015		BAL	Colonisation	5	Spain	Lung tx	F	No	20	2	7	4	1	27	1	
HVH23.1	HVH23.1	26/01/2015		BAL	PCP	5	Spain	haemophilia A	M	Yes	64	2	1	2	2	32	1	
HVH23.2	HVH23.2	29/01/2015		BAL	PCP	5	Spain	haemophilia A	M	Yes	64	2	1	3	2	22	1	
HVH24	HVH24	30/01/2015		BAL	Colonisation	5	Spain	Lung tx	M	No	53	2	1	3	1	21	1	
HVH25	HVH25	4/02/2015		BAL	Colonisation	5	Spain	diffuse scleroderma	F	No	38	2	1	3	1	21	1	
HVH26	HVH26	4/02/2015		BAL	Colonisation	5	Spain	AML	M	No	18	2	1	3	2	22	1	
HVH3	HVH3	25/04/2014		BAL	Colonisation	5	Spain	ALL		No	60	2	1	3	2	22	1	
HVH30	HVH30	25/12/2015		BAL	Colonisation	5	Spain	Lung cancer	M	No	65	2	1	5	1	37	1	
HVH31	HVH31	30/03/2016		BAL	PCP	5	Spain	Liver tx	M	Yes	78	2	1	5	2	38	1	
HVH32	HVH32	4/04/2016		OW	PCP	5	Spain	Lung cancer	M	Yes	55	2	1	5	2	38	1	

HVH35	HVH35	17/05/2016		OW	PCP	5	Spain	Lung cancer	M	Yes	65	2	4	2	2	5	1	
HVH36.1	HVH36.1	24/05/2016		BAL	PCP	5	Spain	Breast Cancer	F	Yes	58	1	1	3	1	39	1	
HVH36.2	HVH36.2	24/05/2016		BAL	PCP	5	Spain	Breast Cancer	F	Yes	58	1	1	3	3	40	1	
HVH38	HVH38	24/08/2016		OW	PCP	5	Spain	Lung cancer	M	Yes	71	1	1	4	2	3	1	
HVH39	HVH39	24/08/2016		BAL	PCP	5	Spain	Lymphoma	M	Yes	51	1	1	2	1	31	1	
HVH4	HVH4	2/05/2014		BAL	Colonisation	5	Spain	Lymphoma	F	No	75	2	1	2	2	32	1	
HVH40	HVH40	27/09/2016		BAL	PCP	5	Spain	Lymphoma	F	Yes	55	1	1	5	2	41	1	
HVH42.1	HVH42.1	25/11/2016		BAL	Colonisation	5	Spain	Astrocytoma	M	No	51	2	2	3	1	48	1	
HVH42.2	HVH42.2	25/11/2016		BAL	Colonisation	5	Spain	Astrocytoma	M	No	51	2	2	5	1	49	1	
HVH43	HVH43	31/12/2016		OW	PCP	5	Spain	Lymphoma	M	Yes	63	2	1	4	1	23	1	
HVH44	HVH44	3/01/2017		OW	PCP	5	Spain	debut	M	Yes	40	1	1	2	1	31	1	
HVH45	HVH45	5/01/2017		BAL	PCP	5	Spain	debut	M	Yes	56	2	1	2	1	30	1	
HVH46	HVH46	5/01/2017		BAL	PCP	5	Spain	Ataxia-telangiectasia	F	Yes	21	5	2	4	1	24	1	
HVH47	HVH47	2/02/2017		BAL	Colonisation	5	Spain	Lymphoma	F	No	60	2	6	2	1	34	1	
HVH48	HVH48	9/02/2017		OW	PCP	5	Spain	Lymphoma	M	Yes	71	2	1	3	2	22	1	
HVH51	HVH51	16/04/2017		BAL	PCP	5	Spain	HIV-positive, Liver tx, lung cancer	M	Yes	60	2	1	5	1	37	1	
HVH52	HVH52	18/04/2017		BAL	Colonisation	5	Spain	Lymphoma	M	No	68	2	1	3	1	21	1	
HVH6.1	HVH6.1	20/06/2014		BAL	PCP	5	Spain	HIV-positive	M	No	43	2	1	5	1	37	1	
HVH6.2	HVH6.2	20/06/2014		BAL	PCP	5	Spain	HIV-positive	M	No	43	2	1	3	1	21	1	
HVH60	HVH60	25/07/2017		BAL	Colonisation	5	Spain	Leriche syndrome	M	No	53	2	7	2	1	35	1	
HVH61	HVH61	25/07/2017		BAL	PCP	5	Spain		M	Yes	37	1	1	4	1	13	1	
HVH62.1	HVH62.1	25/07/2017		OW	PCP	5	Spain	Lymphoma	M	Yes	52	1	1	4	1	13	1	
HVH62.2	HVH62.2	25/07/2017		OW	PCP	5	Spain	Lymphoma	M	Yes	52	1	4	4	1	43	1	

HVH64.1	HVH64.1	28/07/2017		BAL	PCP	5	Spain	Lymphoma	M	Yes	18	2	1	2	1	30	1	
HVH64.2	HVH64.2	28/07/2017		BAL	PCP	5	Spain	Lymphoma	M	Yes	18	2	1	3	1	21	1	
HVH65	HVH65	17/09/2017		BAL	Colonisation	5	Spain	Bladder cancer	M	No	65	1	6	4	1	44	1	
HVH67.1	HVH67.1	28/09/2017		OW	PCP	5	Spain	Colon cancer	M	Yes	63	2	1	3	1	21	1	
HVH67.2	HVH67.2	28/09/2017		OW	PCP	5	Spain	Colon cancer	M	Yes	63	2	1	5	1	37	1	
HVH68	HVH68	3/11/2017		BAL	PCP	5	Spain	myelodysplastic syndrome	F	Yes	53	1	4	2	2	17	1	
HVH69	HVH69	18/11/2017		OW	PCP	5	Spain	Breast Cancer	F	Yes	59	1	4	2	1	33	1	
HVH7.1	HVH7.1	20/06/2014		BAL	PCP	5	Spain	Thymus cancer	M	No	57	2	1	5	1	37	1	
HVH7.2	HVH7.2	20/06/2014		BAL	PCP	5	Spain	Thymus cancer	M	No	57	2	1	3	1	21	1	
HVH70	HVH70	27/11/2017		OW	Colonisation	5	Spain	Lung cancer	F	No	48	1	1	5	2	41	1	
HVH71	HVH71	4/12/2017		OW	Colonisation	5	Spain	Lung cancer	M	No	39	1	1	4	1	13	1	
HVH73	HVH73	8/01/2018		BAL	PCP	5	Spain	CLL	F	Yes	65	2	8	2	1	36	1	
HVH76	HVH76	27/02/2018		BAL	Colonisation	5	Spain	Pulmonary nodule	F	No	58	2	1	5	2	38	1	
HVH78	HVH78	30/03/2018		BAL	PCP	5	Spain	Lung cancer	M	Yes	54	1	2	4	2	42	1	
HVH79	HVH79	6/04/2018		BAL	PCP	5	Spain	HIV-positive	M	Yes	82	1	1	4	2	3	1	
HVH8.1	HVH8.1	25/06/2014		BAL	Colonisation	5	Spain	AML	M	No	17	2	1	5	1	37	1	
HVH8.2	HVH8.2	25/06/2014		BAL	Colonisation	5	Spain	AML	M	No	17	2	1	3	1	21	1	
HVH82.1	HVH82.1	25/06/2018		OW	PCP	5	Spain	Lung cancer	M	Yes	74	1	1	4	1	13	1	
HVH82.2	HVH82.2	25/06/2018		OW	PCP	5	Spain	Lung cancer	M	Yes	74	1	6	4	1	44	1	
HVH85	HVH85	27/07/2018		BAL	Colonisation	5	Spain	Liver cancer + ALL	F	No	57	1	7	4	1	28	1	
HVH9.1	HVH9.1	26/06/2014		BAL	Colonisation	5	Spain	Colon cancer	F	No	71	2	1	5	1	37	1	

HVH9.2	HVH9.2	26/06/2014		BAL	Colonisation	5	Spain	Colon cancer	F	No	71	2	1	3	1	21	1	
NZ1	13E3S4	Nov-17	2018	IS	PCP	6	New Zealand			Yes		1	1	7	2	20	1 v3	N-P
NZ2	13DQ30	Nov-17	2018	IS	PCP	6	New Zealand			Yes		1	1	4	1	13	1	N-P

Supplier

1 = A/Prof. Sharon CA Chen, Westmead Hospital/NSW Pathology, Sydney, Westmead, NSW, Australia
2 = Sebastian van Hall, Royal Prince Alfred Hospital, Sydney, Campertown, NSW, Australia
3 = Rosely M Zancope-Oliveira, INI, FIOCRUZ, Rio de Janeiro, Brazil
4 = Sergio L. Vargas, Instituto de Ciencias Biomédicas (ICBM), Facultad de Medicina, Universidad de Chile, Santiago de Chile, Chile
5 = M. Teresa Martin-Gomez, Hospital Universitari Vall d'Hebron, Barcelona, Spain
6 = Wendy McKinney, Auckland City Hospital, Auckland, New Zealand

Note

N-P = Amplification not performed
N-A = No PCR amplification
N-S = Successful amplification but no sequencing
Empty field = not done or no information

Appendix 2

List of Sequences Generated

***β-Tubulin* (*TUB*) locus**

>Allele Type 1
TCATTAGGTGGTGGAACGGGTGCAGGCATGGGAACTTTGCTAATTTCGAAAATTCGAG
AGGAATATCCGGATCGGATGATGGCAACGTTTTCAGTGGTTCCCTCACCAAAAGTTTCC
GATACAGTTGTAGAGCCATATAATGCAACATTATCAGTGCATGTGTGTTTTTAAGCCAT
TTTTAGAATGTATATTAATGAGGAGGGGTAGCAATTAGTTGAAAATTCCGATGAAACA
TTCTGTATCGACAATGAAGCATTATATGATATTTGTATGCGTACATTAAAATTGCCG

>Allele Type 2
TCATTAGGTGGTGGAACGGGTGCAGGCATGGGAACTTTGCTAATTTCGAAAATTCGAG
AGGAATATCCGGATCGGATGATGGCAACGTTTTCAGTGGTTCCCTCACCAAAAGTTTCC
GATACAGTTGTAGAGCCATATAATGCAACATTATCAGTGCATGTGTGTTTTTAAGCCAT
TTTTAGAATGTATATTAATGAGGAGGGGTAGCAATTAGTTGAAAATTCCGATGAAACA
TTCTGTATCGACAATGAAGCATTATATGATATTTGTATGCGTACATTGAAATTGCCG

>Allele Type 3
TCATTAGGTGGTGGAACGGGTGCAGGCATGGGAACTTTGCTAATTTCGAAAATTCGAG
AGGAATATCCGGATCGGATGATGGCAACGTTTTCAGTGGTTCCCTCACCAAAAGTTTCC
GATACAGTTGTAGAGCCATATAATGCAACATTATCAGTGCATGTGTGTTTTTAAGCCAT
TTTTAGAATGTATATTAATGAGGAGGGGTAGCAATTAGTTGAAAATTCCGATGAAACA
TTCTGTATCGACAATGAAGCATTATATGATATTTGTATGCGTACATTTAAATTGCCG

>Allele Type 4
TCATTAGGTGGTGGAACGGGTGCAGGCATGGGAACTTTGCTAATTTCGAAAATTCGAG
AGGAATATCCGGATCGGATGATGGCAACGTTTTCAGTGGTTCCCTCACCAAAAGTTTCC
GATACAGTTGTAGAGCCATATAATGCAACATTATCAGTGCATGTGTGTTTTTAAGCCAT
TTTTAGAATGTATATTAATGAGGAGGGGTAGCAATTAGTTGAAAATTCCGATGAAACA
TTCTGTATCGACAATGAAGCATTATATGATATTTGTATGCGTACATATAAATTGCCG

>Allele Type 5
TCATTAGGTGGTGGAACGGGTGCAGGCATGGGAACTTTGCTAATTTCGAAAATTCGAG
AGGAATATCCGGATCGGATGATGGCAACATTTTCAGTGGTTCCCTCACCAAAAGTTTCC
GATACAGTTGTAGAGCCATATAATGCAACATTATCAGTGCATGTGTGTTTTTAAGCCAT
TTTTAGAATGTATATTAATGAGGAGGGGTAGCAATTAGTTGAAAATTCCGATGAAACA
TTCTGTATCGACAATGAAGCATTATATGATATTTGTATGCGTACATTAAAATTGCCG

***Cytochrome b* (*CYB*) locus**

>Allele Type 1
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 2
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTTTTGTGG

>Allele Type 4
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTTTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 5
GGTGTAGTTATTTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTTG
GTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTGG
ATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCTG
AATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCTC
ATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCAG
ATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTTT
ATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATTA
TATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTACC
TTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAGC
TGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 6
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 7
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT

GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTTA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 8
GGTGTAGTTATTTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTTG
GTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTGG
ATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCTG
AATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCTC
ATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 9
GGTGTAGTTATTTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTTG
GTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTGG
ATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCTG
AATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCTC
ATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTATTGGTAATTCAG
ATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTTT
ATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATTA
TATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTACC
TTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAGC
TGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 10
GGTGTAGTTATTTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTTG
GTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTGG
ATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCTG
AATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCTC
ATTTAATCTCTTTACATGTTCATGGAAGTAGTAATCCTCTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTTTTGTGG

>Allele Type 11
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTTTGGGTGTTACTGGTAATTCA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT

ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTCTTGTGG

>Allele Type 12
GGTGTAGTTATTTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTTG
GTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTGG
ATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCTG
AATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCTC
ATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTTTGGGTGTTACTGGTAATTTAG
ATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTTT
ATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATTA
TATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTACC
TTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAGC
TGCTATTCTTATTCTTTTTGTTTTACCTTTTGTGG

>Allele Type 13
GGTGTAGTTATCTTCTTAATTATGATTGTTACTGCTTTCTTGGGATATGTTCTGCCTTTT
GGTCAAATGTCATTGTGGGGAGCGACTGTTATTACTAATTTGATGTCTGCTATACCTTG
GATTGGTAATGATATTGTGAATTTTATTTGGGGTGGGTTCTCTGTTAATCATGCTACTCT
GAATTGATTCTTCTCTTTACATTATTTATTGCCTTTTGTTTTATTGGCTTTAGTTGTTGCT
CATTTAATTTCTTTACATGTTCATGGAAGTAGTAATCCTTTGGGTGTTACTGGTAATTTA
GATCGTCTGCCTTTCCATCCCTATTTCTCATTTAAAGATTTAGTTACTGTTTTTTTATTTT
TATTAGCTTTATCTTTCTTTGTGTTTTATGCTCCTAATGTCTTGGGACATAGTGATAATT
ATATTATGGCTAATCCTATGGCTACTCCTCCAAGTATTGTTCCTGAATGGTATCTTTTAC
CTTTCTATGCAATCTTGTGATCTATTTCGAATAAATTATTTGGAGTTGTGGCTATGTTAG
CTGCTATTCTTATTCTTTTTGTTTTACCTTTTGTGG

***Mitochondrial 26S* (*mt26S*) locus**

>Allele Type 1
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTCAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAGACA
GTTA

>Allele Type 2
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTTAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 3
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTCAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG

TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 4
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTAAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 5
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTCAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCTAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 6
TTGTGGTAAGTAGTGAAATACAAATCCGGACTAGGATATAGCTGGTTTTCCTGCGAAA
ATTGTTTTGGCAAATTGTTTATTCCTCTTAAAAATAGTAGGTATAGCACTGAATATCTC
GAGGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAA
TAAATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAAC
AGTAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAA
CAGTTA

>Allele Type 7
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTCAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAATCAG
TTA

>Allele Type 8
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTTAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCCAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAATCAG
TTA

>Allele Type 9
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTCAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCGAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 10
TTGTGGTAAGTAGTGAAATACAAATCGGACTAGGATATAGCTGGTTTTCTGCGAAAAT
TGTTTTGGCAAATTGTTTATTCCTCTAAAAAATAGTAGGTATAGCACTGAATATCTCGA
GGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTTCTTAAAATA
AATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCCCAGAACAG
TAATTAAAGCTCCCGAATTAATATTAAGTGAAATAAAAGTTGTTGGATATCTAAAACA
GTTA

>Allele Type 11
TTGTGGTAAGTAGTGAAATACAAATC-GGACTAGGATATAGCTGGTTTTC-
TGCGAAAATTGTTTTGGCAAATTGTTTATTCCTCTAAAAAATAGTAGGTATAGCACTGA
ATATCTCGAGGGAGTATGAAAATATTTATCTCAGATATTTAATCTCAAAATAACTATTT
CTTAAAATAAATAATCAGACTATGTGCGATAAGGTAGATAGTCGAAAGGGAAACAGCC
CAGAACAGTAATTAAAGCTCCCTAATTAATATTAAGTGAAATAAAAGTTGTTGGATAT
CTAAAACAGTTA

***Superoxide dismutase* (*SOD*) locus**

>Allele Type 1
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGCCCTTTTTTAA
AATTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCT
TTGGAAAAATATAATGAATATGATTCTTCTGTGGATTTAGCAACTCGTATGAATCTTTT
AACATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTT
GAATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAA
AAGAAGGTGGAGGACAAGTTATTGATGGGCCTTTAGTTGATGCAATTAAAAAGGAATG
GGGAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAA
GTGGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACG
ATGGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTA
CTCAAGGCAAAGTTATTCTTGGAATAGTAAAGTTACTTTATATTGTTTTATAAATAATT
AATTGTTTT

>Allele Type 2
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGTCCTTTTTTAA
AATTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCT
TTGGAAAAATATAATGAATATGATTCTTCTGTGGACTTAGCAACTCGTATGAATCTTTT
AACATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTT
GAATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAA
AAGAAGGTGGAGGACAAGTTATTGATGGGCCTTTAGTTGATGCAATTAAAAAGGAATG
GGGAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAA
GTGGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACG
ATGGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTA
CTCAAGGCAAAGTTATTCTTGGAATAGTAAAGTTACTTTATATTGTTTTATAAATAATT
AATTGTTTT

>Allele Type 3
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGTCCTTTTTTAA
AATTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCT
TTGGAAAAATATAATGAATATGATTCTTCTGTGGATTTAGCAACTCGTATGAATCTTTT
AACATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTT
GAATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAA
AAGAAGGTGGAGGACAAGTTATTGATGGGCCTTTAGTTGATGCAATTAAAAAGGAATG

GGGAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAA
GTGGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACG
ATGGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTA
CTCAAGGCAAAGTTATTCTTGGAATAGTAAAGTTACTTTATATTGTTTTATAAATAATT
AATTGTTTT

>Allele Type 4
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGCTTTTTTAAAA
TTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCTTT
GGAAAAATATAATGAATATGATTCTTCTGTGGATTTAGCAACTCGTATGAATCTTTTAA
CATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTTGA
ATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAAAA
GAAGGTGGAGGACAAGTTATTGATGGGCCTTTAGTTGATGCAATTAAAAAGGAATGGG
GAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAAGT
GGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACGAT
GGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTACT
CAAGGCAAAGTTATTCTTGGAATAGTAAAGTTACTTTATATTGTTTTATAAATAATTAA
TTGTTTT

>Allele Type 5
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGTCCTTTTTTAA
AATTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCT
TTGGAAAAATATAATGAATATGATTCTTCTGTGGACTTAGCAACTCGTATGAATCTTTT
AACATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTT
GAATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAA
AAGAAGGTGGAGGACAAGTTATTGATGGGCCTTTAGTTGATGCAATTAAAAAGGAATG
GGGAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAA
GTGGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACG
ATGGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTA
CTCAAGGCAAAGTTATTCTTGGAATAGTAAAGCTACTTTATATTGTTTTATAAATAATT
AATTGTTTT

>Allele Type 6
TTATCTTTCTCATGATTTGCTTGAGGTAAATACTTTTTCTTTGTTTAAAGCCCTTTTTTAA
AATTATAGCTTCATTATAACAAACATCACCGTGCTTACGTAACAAATTTTAATTTAGCT
TTGGAAAAATATAATGAATATGATTCTTCTGTGGATTTAGCAACTCGTATGAATCTTTT
AACATCTATTAAGTTTCATGGTGGTGGTAGGTATAGGAAAGATAAGAACTATTGATTT
GAATATTTTTTATAGGTCATATTAATCATTCTTTATATTGGGAAAGCCTTCTTCCACCAA
AAGAAGGTGGAGGACAAGTTATTGATGGACCTTTAGTTGATGCAATTAAAAAGGAATG
GGGAAGTGTTGACCAATTCATTCGTACATTTAATACACATTTGTCTGGGATTCAAGGAA
GTGGGTGGTGTTGGCTCGTAAAAATACCTTCAAGTCGACAACTTTTTATTCAAACAACG
ATGGTACTTTTCTTCTTATACTCTTTAGTGTCTGATTTGAATAGAATCAAGATCTTGTTA
CTCAAGGCAAAGTTATTCTTGGAATAGTAAAGTTACTTTATATTGTTTTATAAATAATT
AATTGTTTT

***Dihydropteroate Synthases* (*DHPS*)**

>Allele Type 1 (also known as wild type)
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGAGATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGGCCGTGTAGACTGCCCACC
AATATCAATTATCGTCGCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAAT
CATATGAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

>Allele Type 1 Variant 1
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGACATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGGCCGTGTAGACTGCCCACC
AATATCAATTATCGTCGCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAAT
CATATGAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

>Allele Type 1 Variant 2
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGAGATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGGCCGTGTAGACTGCCCACC
AATATCAATTATCGTCGCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAAT
CATATCAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

>Allele Type 1 Variant 3
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGAGATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGGCCGTGTAGACTGCCCACC
AATATCAATTATCGTCGCCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAA
TCATATGAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

>Allele Type 3
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGAGATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGGCCGTGCAGACTGCCCACC
AATATCAATTATCGTCGCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAAT
CATATGAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

>Allele Type 4
ACTTTCAACTTGGCAACCACATTAAGCATTTTTGGATCATACCTTCCCCCACTTATATCA
TTAACAAGACTAGCACCAGCCTTAATTGCTTGTTCTGCAACCTCAGAACGAAAAGTATC
TACACTTACTAAAATATCAGGATATACTTTTAAGAGATATTTTATAGCAGGAATAACTC
GAGAAATCTCTTCCTCTATAGAAACAACATGTGAACCAGACCGTGCAGACTGCCCACC
AATATCAATTATCGTCGCCCCTGCATTTATAAAATTCTCCACATCCATTAATATAGAAT
CATATGAATGAACACCCCCATCGAAAAAAGAATCAGGAGTAAGATTTAAAATGGCC

Internal Transcribed Spacer (ITS)

>Allele Type 2
ATTCAGCTTAAACACATCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTTT
TTGTTTGGCGAGGAGCTGGCTTTTTTTGCTTGCCTCGCCAAAGGTGTTTATTTTTAAAATT
TTAAATTGAATTTCAGTTTTAGAATTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTC
TCGCGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGT
GAATCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGT
TTGAGCGTTATTTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAA
ATATAATTAGAATAAAATATTTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTT
TTCTTTGCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTATAGATA
CAAGATTT

>Allele Type 7
ATTCAGCTTAAACACTTCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTT
TGTTTGGCGAGGAGCTGGCTTTTTTTGCTTGCCTCGCCAAAGGTGTTTATTTTTAAAATTT
TAAATTGAATTTCAGTTTTAGAATTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTCT
CGCGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGTG
AATCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGTT
TGAGCGTTATTTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAAA
TATAATTAGAATAAAATAATTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTTT
TCTTTGCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTTATAAATA
CAAGAATT

>Allele Type 8
ATTCAGCTTAAACACTTCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTT
TTGTTTGGCGAGGAGCTGGCTTTTTTTGCTTGCCTCGCCAAAGGTGTTTATTTTTAAAATT
TTAAATTGAATTTCAGTTTTAGAATTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTC
TCGCGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGT
GAATCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGT
TTGAGCGTTATTTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAA
ATATAATTAGAATAAAATAATTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTTT
TCTTTGCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTTATAAAT
ACAAGAATT

>Allele Type 9
ATTCAGCTTAAACACTTCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTT
TGTTTGGCGAGGAGCTGGCTTTTTTTGCTTGCCTCGCCAAAGGTGTTTTTTTTAAAATTTTA
AATTGAATTTCAGTTTTAGAATTTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTCTCG
CGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGTGAA
TCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGTTTG
AGCGTTATTTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAAATA
TAATTAGAATAAAATAATTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTTTTC
TTTGCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTTATAAATACA
AGAATTT

>Allele Type 10
ATTCAGCTTAAACACTTCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTT
TTGTTTGGCGAGGAGCTGGCTTTTTTTGCTTGCCTCGCCAAAGGTGTTTATTTTTAAAATT
TTAAATTGAATTTCAGTTTTAGAATTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTC
TCGCGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGT
GAATCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGT

TTGAGCGTTATTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAAA
TATAATTAGAATAAAATATTTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTTT
TCTTTGCCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTTATATAG
ATACAAGAATTT

>Allele Type 11
ATTCAGCTTAAACACTTCCCTAGTGTTTTAGCATTTTTCAAACATCTGTGAATTTTTTTT
TTGTTTGGCGAGGAGCTGGCTTTTTTGCTTGCCTCGCCAAAGGTGTTTATTTTTAAAATT
TTAAATTGAATTTCAGTTTTAGAATTTTTTAAAAACTTTCAACAATGGATCTCTTGGCTC
TCGCGTCGATGAAGAACGTGGCAAAATGCGATAAGTAGTGTGAATTGCAGAATTTAGT
GAATCATCGAATTTTTGAACGCATCTTGCGCTCCTTAGTATTCTAGGGAGCATGCCTGT
TTGAGCGTTATTTTAAGTTCCTTTTTTCAAGCAGAAAAAAGGGGATTGGGCTTTGCAAA
TATAATTAGAATAAAATATTTATATGCATGCTAGTCTGAAATTCAAAAGTAGCTTTTTT
TCTTTGCCTAGTGTCGTAAAAATTCGCTGGGAAAGAAGGAAAAAAGCTTTTATATAGA
TACAAGAATTT

Appendix 3

List of Chemicals, Kits and Reagents

AP3.1 List of Chemicals, Kits and Reagents

Chemicals, Kits and Reagents	Supplier
1Kb Plus DNA ladder	Invitrogen Australia Pty. Ltd., Mulgrave, VIC, Australia
Agarose	Amresco, Solon, Ohio, USA
10 x NH4 PCR Buffer	Bioline Australia Pty. Ltd., Alexandria, NSW, Australia
BioTaq™ DNA Polymerase	Bioline Australia Pty. Ltd. Alexandria, NSW, Australia
Boric Acid	Research Oraganics Inc., Cleaveland, OH, USA
Bromophenol Blue	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Deoxynucleotide (dNTPs)	Bioline Australia Pty. Ltd., Alexandria, NSW, Australia
Ethanol	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Ethidium Bromide (EtBr)	Bio-Rad Laboratories Pty., Ltd., Gladesville, NSW, Australia
Ethylenediaminetetraacetic acid (EDTA)	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Glycerol	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
HL-SAN DNase	ArcticZymes Technologies ASA, Norway
Liquid nitrogen	BOC, Linde plc, Ryde, NSW, Australia
Magnesium chloride (MgCl2)	Bioline Australia Pty. Ltd., Alexandria, NSW, Australia
Nuclease-free water	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Sodium chloride (NaCl)	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Sputum DNA Isolation Kit	Norgen Biotek Corp., Millennium Science Australia Pty Ltd., Mulgrave, VIC, Australia
Phosphate-buffered saline (PBS)	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Primers and Oligonucleotides	Bioline Australia Pty. Ltd., Alexandria, NSW, Australia
Proteinase K	New England Biolabs Inc., Ipswich, MA, USA
Saponin	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
Tris-HCL	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia
QIAamp DNA Blood Kit	QIAGEN Pty Ltd., Chadstone, VIC, Australia
QIAamp DNA Mini Kit	QIAGEN Pty Ltd., Chadstone, VIC, Australia
Quick-DNA Fungal/Bacterial Miniprep Kit	Zymo Research, Irvine, CA, USA
Ultra Deep Microbiome PrepEnrichment Kit	Molzym GmbH & Co. KG., Bremen, Germany
Xylene Cyanol FF	Sigma – Aldrich Pty. Ltd., Castle Hill, NSW, Australia

AP3.2 List of Computer software and Databases used in this study

Computer Software	Producer
BioEdit 7.1.9.	Ibis Biosciences, CA, USA
BioloMICS ver. 21.07.9.324	BioAware, Hannut, Belgium
BLAST	Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.
DnaSP version 5.10.01	University of Barcelona, Barcelona, Spain
Fungal MLST Database	http://mlst.mycologylab.org NCBI, Maryland, U.S. Inc., Oxford, UK
Genbank Database	https://www.ncbi.nlm.nih.gov/genbank/
LIAN version 3.5	Max Planck Institute Jena, Jena, Germany and University of Chicago, Chicago, USA
MEGA7 Database	Tokyo Metropolitan University, Tokyo, Japan
Microsoft Excel	Microsoft, Redmond, WA, USA
RaxmlGUI version 1.1	Research Institute Senckenberg, Frankfurt am Main, Germany
Sequencher™	Sequencher ™ Gene Codes Corporation, Ann Arbor, MI, USA
Splitstree version 4.10	Tübingen University, Tübingen, Germany
What's In My Pot (WIMP) Database	Oxford Nanopore Technologies

AP3.3 List of Equipment used in this study

Equipment	Suppliers
Automatic pipettes (P2, P10, P200, P1000)	Pipetman, Gilson Medical Electronics, Villiers-le-Bel, France
Bench top microfuge (5415C, 5415D)	Eppendorf-Netherler-Hinz, GmbH, Hamburg, Germany
Canon photo camera PowerShot G3	Canon Lane Cove Road, Macquarie Park NSW, Australia
DS-11 Series Spectrophotometer	DeNovix Inc., Wilmington, NC, USA
Dry block heater	IKA Works, Wilmington, NC, USA
Electrophoresis chamber	Bio-Rad Laboratories Pty., Ltd., Gladesville, NSW, Australia
HulaMixer	Thermo Fisher Scientific, Waltham, MA, USA
Magnetic stirrer	IKA Works, Wilmington, NC, USA
Nanopore MinION	Oxford Nanopore Technologies, Inc., Oxford, UK
Thermal cycler (LabCycler D-37085)	SensoQuest, Göttingen, Germany
UV transilluminator TM40E (λ=302nm)	UniEquip, Martinreid, Munich, Germany
Vortex (MS1 minishaker)	IKA Works, Wilmingtom, NC, USA

Appendix 4

Home page of the Database for the *Pneumocystis jirovecii* MLST Scheme (Phipps *et al.* 2011)

Home page of the Database for the *Pneumocystis jirovecii* MLST Scheme (Phipps *et al.* 2011)

https://mlst.mycologylab.org/page/PJPhipp2011

Appendix 5

Home page of the Database for the *Pneumocystis jirovecii* MLST Scheme (Pasic *et al.* 2020)

Home page of the Database for the *Pneumocystis jirovecii* MLST Scheme (Pasic *et al.* 2020)

https://mlst.mycologylab.org/page/PJPasic2020