Comparison of MultiLocus Sequence Typing (MLST) and Microsatellite Length Polymorphism (MLP) for Pneumocystis jirovecii genotyping

Maud Gits-Muselli, Pascal Campagne, Marie Desnos-Ollivier, Patrice Le Pape, Stéphane Bretagne, Florent Morio, Alexandre Alanio

Computational and Structural Biotechnology Journal 18 (2020) 2890–2896

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Comparison of MultiLocus Sequence Typing (MLST) and Microsatellite Length Polymorphism (MLP) for *Pneumocystis jirovecii* genotyping

Maud Gits-Muselli [a,b,c,1], Pascal Campagne [d,1], Marie Desnos-Ollivier [c], Patrice Le Pape [e,g], Stéphane Bretagne [a,b,d], Florent Morio [e,f], Alexandre Alanio [a,b,c,*]

[a] *Laboratoire de Parasitologie-Mycologie; AP-HP, Groupe Hospitalier Saint-Louis-Lariboisière-Fernand-Widal, Paris, France*
[b] *Université de Paris, Sorbonne Paris Cité, Paris, France*
[c] *Institut Pasteur, CNRS, unité de Mycologie Moléculaire, Centre National de référence Mycoses invasives et Antifongiques (CNRMA), UMR2000, Paris, France*
[d] *Hub of Bioinformatics and Biostatistics – Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France*
[e] *Laboratoire de Parasitologie-Mycologie, Institut de Biologie, CHU Nantes, Nantes, France*
[f] *Département de Parasitologie et Mycologie Médicale, EA1155 IICiMed, Institut de Recherche en Santé 2, Université de Nantes, Nantes Atlantique Universités, Nantes, France*

ARTICLE INFO

Article history:
Received 9 July 2020
Received in revised form 28 September 2020
Accepted 2 October 2020
Available online 15 October 2020

Keywords:
Pneumocystis jirovecii
MultiLocus Sequence Typing
Genotyping
MLP
Haplotypes

ABSTRACT

Pneumocystis jirovecii is an atypical fungus responsible for severe respiratory infections, often reported as local outbreaks in immunocompromised patients. Epidemiology of this infection, and transmission risk emphasises the need for developing genotyping techniques. Currently, two methods have emerged: Multilocus Sequence typing (MLST) and microsatellite length polymorphism (MLP). Here we compare an MLST strategy, including 2 nuclear loci and 2 mitochondrial loci, with an MLP strategy including 6 nuclear markers using 37 clinical PCR-positive respiratory samples from two French hospitals. *Pneumocystis jirovecii* MLST and MLP provided 30 and 35 different genotypes respectively. A higher number of mixed infections was detected using MLP (48.6% vs. 13.5% respectively; $p = 0.002$). Only one MLP marker (STR279) was statistically associated with the geographical origin of samples. Haplotype network inferred using the available genotypes yielded expanded network for MLP, characterized by more mutational steps as compared to MLST, suggesting that the MLP approach is more resolutive to separate genotypes. The correlation between genetic distances calculated based on MLST and MLP was modest with a R^2 value = 0.32 ($p < 0.001$). Finally, both genotyping methods fulfilled important criteria: (i) a discriminatory power from 97.5% to 99.5% and (ii) being quick and convenient genotyping tools. While MLP appeared highly resolutive regarding genotypes mixture within samples, using one genotyping method rather than the other may also depend on the context (i.e., MLST for investigation of suspected clonal outbreaks versus MLP for population structure study) as well as local facilities.

1. Introduction

Pneumocystis jirovecii is an atypical fungus responsible for Pneumocystis pneumonia (PCP), a severe respiratory infection in immuno-compromised patients such as HIV, organ transplant recipients, patients affected by hematological malignancies and patients under immunosuppressive therapies [1]. The incidence of PCP is estimated more than half a million cases every year, worldwide, associated with a mortality of 50% (http://www.gaffi.org/). In France, *P. jirovecii* is also the second most common fungal agent responsible for invasive fungal infection, although its incidence is not precisely known since notification is not mandatory [2]. Nevertheless, the surveillance system conducted by the National Reference Center of Invasive Mycoses and Antifungals, with the participation of 30 voluntary medical centers counted around 900 cases every year, with a three months mortality ranging from 7% to 36% according to the underlying disease, respectively AIDS and hematological malignancy [1].

To better understand the physiopathology and epidemiology of medically important fungi, several of genotyping methods have been developed [3]. For *P. jirovecii*, in the absence of a reliable culture system for this organism, molecular methods targeting the fungus within respiratory specimens are the only way to study *P. jirovecii* transmission [4] and its genetic diversity [5]. To this purpose, different tools have been developed in *P. jirovecii* since the

* Corresponding author at: Molecular Mycology unit, Institut Pasteur, 25 rue du Dr Roux 75724 Paris Cedex 15, France
E-mail address: alexandre.alanio@pasteur.fr (A. Alanio).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2020.10.005

1990 s [3], including Single-Strand Conformation Polymorphism (SSCP) [6], Internal Transcribed Spacer sequencing [3], Single Base Extension technology (SBE) [7], and high-throughput pyrosequencing of specific targets [8]. However, some recent genotyping studies also used Multi Locus Sequence Typing [9] and microsatellite length polymorphism (MLP) [10]. Thus, an MLST approach including two nuclear and two mitochondrial loci with a discriminatory power > 0.99 has been proposed, based on existing loci combination [11]. In parallel, an MLP approach based on six short tandem repeats (STR) DNA markers with a discriminatory power > 0.99 was developed and tested in different cohorts including a cohort of patients originating from 16 European centers [5,10,12].

Comparison of MLST and MLP typing methods have already been performed for medically important fungi such as *Candida albicans* [13,14]. Whereas MLST and MLP strategies yielded similar performances in *C. albicans* [15], MLP was found to be a better option in *A. fumigatus* [16]. We were therefore interested in comparing MLST and MLP performances in *P. jirovecii* genotyping. Using *P. jirovecii* PCR-positive specimens, collected from unrelated patients, in two French hospitals 400 km apart, we aimed at comparing one validated MLST [11] and MLP strategies [10] with respect to (i) their discriminatory power, (ii) their ability to detect mixtures, (iii) their potential to detect geographical clustering, (iv) the structure of their haplotype networks, and (v) the relationship between Sequence types / genotypes.

2. Methods

2.1. Samples selection

Thirty-seven unrelated patients' respiratory samples positive for *P. jirovecii* were included in the study. Nineteen respiratory specimens originated from patients living in Nantes, France, and admitted to Nantes university hospital (henceforth NTS), located 400 km west of Paris and 18 from patients living in Paris area, France, and admitted to Saint-Louis university hospital Paris, (henceforth SLS) were analyzed. The samples were randomly selected among PCR-positive broncho-alveolar-lavage (BAL) fluids using qPCR assays targeting mtLSU using two in-house protocols [17,18]. We selected only samples harboring a high fungal load to enhance the practicability of genotyping methods, (i.e obtaining reliable results with both methods), knowing that MLP methods is limited to high or medium fungal load [10]. All analyses were performed blindly with MLST testing performed in Nantes and MLP testing in Paris. Due to constraints in the sample quantity, genotyping was performed only once. Indeed, performing both MLP and MLST methods required a minimum of 32 μL of extracted DNA samples out of 50 μL available, which prevent performing several times each analysis.

2.2. MLST method

MLST genotyping was performed at the four following loci, two nuclear loci: internal transcribed spacer 1 (ITS1), superoxide dismutase (SOD) and two mitochondrial loci: large subunit of the mitochondrial rRNA gene (mt26S) and cytochrome *b* (CYB) [11]. PCRs were carried out in a 25 μL final volume using Premix Ex Taq (perfect real-time) (Takara Bio, Inc., Otsu, Shiga, Japan) and 5 μL of each DNA extract. The final concentration of each primer was 0.5 μM. The primers used were previously described in Maitte *et al* study [11]. Amplification was conducted on an Applied GeneAmp 9700 (Applied Biosystems, Foster City, CA) under the following conditions: 7 min at 94 °C followed by 35 cycles, including 30 s at 94 °C, 45 s at 60 °C, 30 s at 72 °C, and a final elongation step at 72 °C for 7 min. PCR products were purified and sequenced on a 3130xl genetic analyzer (Applied Biosystems). Nucleotide sequences were analyzed using the SeqScape software (Applied Biosystems). Sequencing of these loci allowed the detection of single nucleotide polymorphism within the selected region, and sequences were compared to the following reference sequences with these Genbank accession numbers: U07220 (*ITS1*), AF146753 (*SOD*), M58605 (*mt26S*) and AF320344 (*CYB*).

Nucleotide polymorphic positions considered for analysis were positions 2, 8–10, 11, 17, 18, 22, 46–47, 70–71, 79 and 111–113 for *ITS1* ; 110, 191 and 215 for *SOD* ; 54–57, 80, 85, 248, 288 for *mt26S* ; 279, 299, 348, 362, 369, 516, 574, 566, 742, 832–833 and 838 for *CYB*. Only sequences with "pure" chromatograms (one peak corresponding to a single nucleotide at every position) or a with only one mixed marker (two peaks observed at a polymorphic position) could lead to Sequence Type (ST) assignation. If alleles were previously described, they were named according to published nomenclature [6,11,19,20].

2.3. MLP method

The six short tandem repeat (STRs) markers were amplified separately by PCR [10]. The forward primers were tagged with fluorophores (FAM, HEX or ATTO565). All PCR reactions were performed on a GeneAmp PCR System 9700 Thermocycler (Applied Biosystems) in a final volume of 20 μL containing 1X Ampli Taq Gold buffer (Life technologies) with 0.25 μM of each primer, 2.5 mM of MgCl2, 0.8 μM of dNTPs, 0.25 UI of Ampli Taq Gold polymerase (Life technologies) and 2 μL of DNA. The reaction consisted of 10 min at 95°, followed by 35 cycles of 30 s at 95 °C (denaturation), 30 s at 56° (primer annealing) and 60 s at 72 °C (extension) followed by a final extension of 10 min at 72 °C. A sample with a known genotype was used in each PCR run as an internal control to measure reproducibility. After amplification, 2 μL of PCR product was prepared for fragment analysis by the addition of 18 μL of formamide (3700 formamide, Life technologies) and 1 μL of Genescan-500 LYZ Size Standard (Life technologies). Capillary electrophoresis was performed with the denaturing polymer POP-7 (Life technologies) in a 50 mm capillary tube at 60 °C. The lengths of the PCR fragments were determined on an ABI 3500 genetic analyzer with ABI Gene mapper v4.1 software (Life technologies). Genotypes were determined if each of the six STRs markers (STR022, STR108, STR138, STR189, STR278, STR279) were pure (i.e only one detectable peak corresponding to one amplicon length, for each STR marker) or if only one out of six markers present additional peaks, (i.e. corresponding to a mixture of amplicons of different lengths for this marker). In samples harboring mixtures for more than one marker, no genotype determination was possible. The typeability of both method was calculated as follow: percentage of determinable ST/genotype divided by total number of typed samples [21].

Each allele was named according to the length of the amplified fragment in base pair (bp). Four of these markers were trinucleotide repeats [STR022 (amplicon size from 138 to 144 bp), STR108 (138 bp), STR138 (163 to 175 bp) and STR279 (175 to 190 bp)] and the other two were di-nucleotide repeats [STR189 (193 to 219 bp) and STR278 (187 to 191 bp)].

Each run contained a positive control specimen, which gave similar STR allele determination in each run.

2.4. Data analysis

Haplotype networks were inferred from samples that included complete STs/genotypes determination in the two methods. This

means that samples presenting at least two mixed markers with MLP or MLST were excluded. Haplotype networks were inferred using the statistical parsimony method described by Templeton, Crandall and Sing (TCS) in 1992 [22] using the *ape* package [23]. TCS approach presents the ability to handle both amplicon length and DNA sequences data [24]. The generation of these networks allows to schematically represent the most likely connections between STs/genotypes based on statistical parsimony approach for every genotyping method. For MLST, we used the concatenation of the DNA sequences of nucleotide polymorphic positions at the four loci to infer the haplotype network. All data analyses were performed using the R software.

Rarefaction curves were computed based on each dataset (MLP and MLST). Their calculation consisted in counting the number of fungal types obtained with a set of genetic markers (i.e., 5 polymorphic MLP markers, and 4 MLST markers) in random subsamples of varied size (from 1 to 35 observations). The curves were obtained by averaging across 5,000 repetitions of the random process. Indeed, one MLP marker (STR108) did not vary in this dataset.

In addition, to ensure a fair comparison of both methods, we estimated rarefaction curves based on the same number of markers in each genotyping strategy: combinations of the four MLP markers among the five with the most allelic varieties (STR22, STR138, STR189, STR278 and STR279) were thus compared to the four MLST markers. The analysis was also performed comparing MLST to the six MLP markers.

We estimated the discriminatory power in both microsatellites markers (as well as in the different combinations of 4 MLP markers) and MLST markers using a Jackknife procedure [25].

2.5. Statistical analysis

For contingency table analysis, we used χ^2 and Fischer's exact test. Median values and interquartile ranges are presented and Kruskal-Wallis test were performed when the distribution of the values was not normal. The correlation between matrices of genetic distances generated with both MLP and MLST data was tested using a Pearson's correlation test. P values below 0.05 were considered as statistically significant.

2.6. Ethics statement

This study was a retrospective non-interventional study with no change in the usual diagnostic procedures. Biological material and clinical data were obtained only for standard diagnostic following physicians' prescriptions. In France, according to the French Health Public Law (CSP Art L1121-1.1), such study dealing with non-human DNA and with no consequences for the management of the patient does not require approval of an ethics committee and is exempted from specific informed consent application.

3. Results

3.1. Selected samples

Thirty-seven specimens were selected (one per patient). Nineteen Patients were HIV-positive patients [11 from NTS (58%), 8 from SLS (44%)] and eighteen non-HIV patients [8 from NTS (42%) and 10 from SLS (56%)]. In NTS, non-HIV patients were 4 (21%) solid-organ transplant recipients and 4 (21%) others various backgrounds, whereas in SLS, 6 (33%) were from hematology-oncology wards, one (6%) was a solid organ transplant recipient and 3 (17%) had others various backgrounds. These differences between hospitals were not statistically significant (p = 0.75).

3.2. Detection of mixed genotypes

The ability to detect mixtures of STs/genotypes in specimens was significantly higher when using the MLP method, than the MLST method, with 18/37 (48.6%) specimens containing mixed genotype using MLP versus 5/37 (13.5%) using MLST (p = 0.002, Table 1). In details, 4/37 specimens (10.8%) were detected with mixed genotypes with both methods, only 1/37 (2.7%) using the MLST, and 14/37 (37.8%) samples using MLP. In total, based on MLP, the most sensitive method to detect mixed infections, 19/37 (51.3%) specimens were classified as pure, 10/19 (52.6%) samples from NTS, and 9/18 (50%) from SLS (Table 1).

The typeability of MLST and MLP methods was 86.5% (32/37 samples) for MLST and 78.3% (29/37 samples) for MLP. These results are presented in Supplementary Table 1.

3.3. Genetic diversity estimated by the number of genotypes

Thirty and 35 genotypes were assigned from MLST and MLP, respectively. Identical STs/genotypes were found in several unrelated samples described in Table 2.

MLST identified ten new allelic combinations, corresponding to new STs as compared to ST previously described in Maitte study [11]. Among alleles obtained with MLST, most of them, 22/26 (84.6%) had been previously described [6,11,19,20]. Only four new alleles were obtained for *ITS1* locus, all observed in samples from Paris. Allele's details are presented in Table 1 and in Supplementary Table 1.

Based on MLP data, we observed that 12/35 (34.3%) (Gt 1, 3, 10, 11, 13, 17, 18, 21, 23, 24, 25, 26) have been previously described in Europe [5] from various geographical origin: France, Spain, Portugal, UK, Holland, Germany and Czech republic. These genotypes corresponded to 7 samples from Nantes and 5 from Paris. The amplicons lengths alleles of the MLP analysis are described in Table 1.

3.4. Microsatellites markers combinations, MLST, rarefaction curves and discriminatory power

When using the four most discriminant MLP markers, we could observe that the discriminatory power (DP) was 0.991 as compared to 0.995 when the 6 markers were included (Table 3). The discriminatory power was 0.975 for the four MLST markers. All DP values obtained with the different combinations are presented in Table 3.

Calculated DP values are in adequacy to the graphic representation of rarefaction curves proposed in Fig. 1.

4. Geographic distribution, medical background and clustering

We observed a significant association between STR279 allele distribution and the geographical origin of the samples i.e. NTS vs. SLS (p = 0.0003). Indeed, for marker STR279, among the five alleles observed, the 178-bp allele was statistically linked to NTS samples (p = 0.0007), whereas the 181-bp allele was statistically linked to SLS samples (p = 0.002). No association between markers/alleles and geographic distribution was observed from the MLST dataset (p = 0.37).

We did not observe any statistical association of a specific allele with a specific medical background for MLP data (p value = 0.76) or for MLST data (p value = 0.83).

4.1. Haplotype networks

The most likely haplotype networks based on 31 MLP and 23 MLST genotypes/STs are presented in Fig. 2. These corresponded

Table 1
Data associated with genotyping according to each typing approach.

Method	Number of genotypes		Pure samples (%)	Mixed samples (%)		Identities of alleles for each locus (ID for MLST and bp for MLP)			
MLST	30		32(86.5%)	5(13.5%)		**SOD**	**ITS1**	**mt26S**	**CytB**
						n = 3	**n = 13**	**n = 4**	**n = 6**
						SOD1	A3	allele 2	cyb1
						SOD2	A4	allele 3	cyb2
						SOD5	A5	allele 7	cyb5
							B	allele 8	cyb6
							B1		cyb7
							B2		cyb8
							B3		
							B4		
							B6		
							4 new alleles		
MLP	35	19(51.4%)	18(48.6%)	STR022	STR108	STR138	STR189	STR278	STR279
				n = 3	n = 1	n = 5	n = 6	n = 3	n = 5
				138	138	163	193	187	175
				141		166	195	189	178
				144		169	205	191	181
						172	207		184
						175	217		190
							219		

MLST, multilocus sequence typing; MLP, microsatellite length polymorphism; bp, base pair.

Table 2
Details on identical genotypes.

Method	Numbers of identical Sequence Types/ genotypes	Distribution of the identical Sequence Types/ genotypes
MLST	2	Sequence Type 1 (2 samples from NTS) Sequence Type 2 (2 samples: 1 from SLS, 1 from NTS)
MLP	4	Genotype 7 (2 samples from NTS) Genotype 10 (3 samples from SLS) Genotype 22 (2 samples from NTS) Genotype 30 (2 samples: one from SLS, one from NTS)

MLST, multilocus sequence typing; MLP, microsatellite length polymorphism

to 25 samples (13 from NTS, 12 from SLS), this samples are identified with an asterisk in the Supplementary Table 1.

Haplotype network corresponding to MLP data exhibited more mutational steps as compared with MLST (Fig. 2). The total number of mutational steps between haplotypes was 81 for the MLP and 30 for MLST, with 32 and 19 connections, respectively. The median number of mutational steps between two haplotypes was 2 [2; 4] for the MLP network and 2 [1; 2] for the MLST network. The distribution of the number of mutational steps needed to infer these two networks was significantly different (p = 0.0001).

4.2. Genetic distances

The correlation between MLST and MLP genetic distances (Manhattan distances) was modest with a R^2 value = 0.32 when using a Pearson correlation test ($p = 3.4\ 10^{-15}$).

5. Discussion

To our knowledge, this study is the first to compare MLST with MLP for genotyping *P. jirovecii* from clinical materials. We aimed at comparing the number of STs/genotypes obtained with each method based on 37 clinical samples, their ability to detect mixed genotypes, the geographical distribution of alleles, the haplotype networks (connection between genotypes) and the correlation of the two methods. The main result of the comparison is a modest correlation between the two methods (R^2 = 0.32) when comparing the distance matrices.

A previous study aiming at comparing MLST and MLP approaches for *Candida albicans* genotyping presented a good correlation with a R^2 at 0.79 [13]. The poor correlation observed in our study could be due to the localization of the loci selected with MLST scheme including both nuclear loci and mitochondrial loci as compared to six nuclear loci with MLP. For others fungi of medical interest, such as *Candida albicans*, *Aspergillus fumigatus*,

Table 3
Discriminatory power in MLP and MLST markers.

	MLP loci					Discriminatory power	
marker	**STR022**	**STR138**	**STR189**	**STR278**	**STR279**	estimate	95%-CI
comb1 MLP	×	×	×	×		0.984	[0.984, 1]
comb2 MLP	×	×	×		×	0.989	[0.989, 1]
comb3 MLP	×	×		×	×	0.959	[0.959, 0.989]
comb4 MLP	×		×	×	×	0.975	[0.975, 0.993]
comb5 MLP		×	×	×	×	**0.991**	**[0.991, 1]**
all MLP	×	×	×	×	×	**0.995**	**[0.995, 1]**
MLST	nr	nr	nr	nr	nr	0.975	[0.975, 0.993]

The discriminatory power was computed based on the full MLST and MLP datasets but also on the different combinations of four microsatellites markers (denoted by different combinations of crosses in the Table), to allow a direct comparison between MLST and MLP markers. The 95%-Confidence Intervals were estimated by using a Jackknife procedure. STR (Short tandem Repeat).

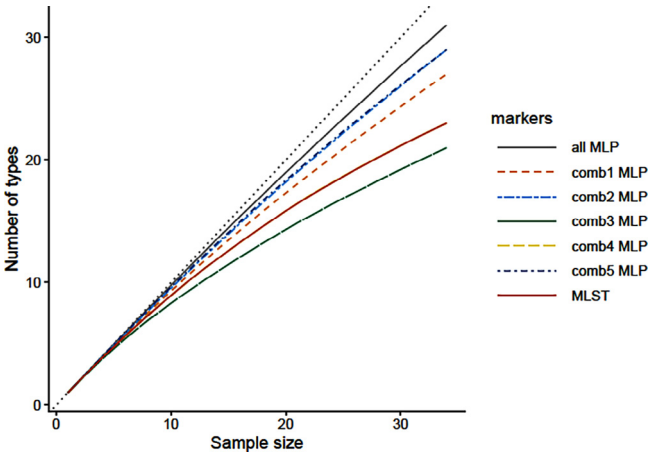

Fig. 1. Rarefaction curves of MLP and MLST markers. The curves were computed based on the full MLST and MLP datasets but also on the different combinations of four microsatellites markers (chosen among STR22, STR138, STR189, STR278 and STR279), to allow a direct comparison between MLST and MLP markers. The different curves represent the different datasets as well as the different combinations of MLP markers. The dotted grey line (with a slope of 1) corresponds to the idealized case where adding one observation in the sample would necessarily result in one additional fungal type.

or *Cryptococcus neoformans*, MLST approaches are based only on nuclear markers.

For *P. jirovecii*, in order to guarantee the best success of both genotyping approaches, we were forced to select only samples with a medium to high fungal load. This is a limitation of this study, as the MLP approach uses single-copy nuclear genes and therefore may not succeed in samples with low fungal load. This limitation is also true for the nuclear loci tested in the MLST design (i.e. the ITS and SOD loci in our study). Therefore, we cannot exclude a potential bias in the determination of STs/genotypes, as previously observed in the study by Alanio and colleagues [26], where certain genotypes were correlated to samples harboring a low fungal load.

The availability of cultures for others fungal pathogen allows to work on a large amount of fungal DNA, which is not possible so far for *P. jirovecii*, and may have encourage the use of repeated mitochondrial markers, more prone to amplification successes [27]. The diversity of MLST alleles found in mitochondrial loci are not negligible (4 alleles for mt26S and 6 for CytB). This should lead to misinterpretation of the diversity and the genetic properties of *P. jirovecii* genome evolution since evolution rates of mitochondrial DNA and nuclear DNA are likely to differ [28]. Moreover, considering our lack of knowledge on mitochondrial inheritance in *P. jirovecii*, and the diversity of inheritance mechanisms described for fungi we should prefer genotyping approaches focused on nuclear genome loci [29]. Fungal phylogenetic relatives of *P. jirovecii* such as *Saccharomyces cerevisiae* use biparental inheritance for mitochondria and recombination of mitochondrial genomes are described to occur during sporulation process [30]. In addition, for *P. jirovecii*, heteroplasmy of mitochondrial DNA have been suggested [8] and mixture of mitochondrial DNA could be therefore hypothesized [27]. These observations add more complexity and uncertainty in the interpretation of sequences of mitochondrial DNA. Therefore, we advocate to use only nuclear markers in the MLST design in genotyping studies.

Based on previous published studies, both MLST and MLP are known to provide a high discriminatory power ≥ 0.99 [10,11] suggesting that both are applicable for genotyping *P. jirovecii*. In our study, we first ensured that samples were independent in choosing PCP samples originated from two different medical centers and from patients with different underlying conditions. When analyzed with MLST and MLP, the 37 selected samples provided a closed number of different STs/genotypes (30 and 35, respectively). Acknowledging the limited numbers of samples (n = 37), the discriminatory power calculated with the two methods could be biased [21]. Here, MLP calculated discriminatory power depends on the marker combination and could therefore present a lower or higher value when compared to MLST. On this limited dataset, STR_108 marker was excluded of the analysis considering that only one allele (amplicon length 138 bp) was observed in the dataset. This was not a surprise, acknowledging that in the European study using MLP, performed on a larger dataset (249 samples), a limited number of variant were described for this marker (135, 138 and 141 bp), with predominance of the 138 bp allele [5].

Our previous studies using MLP were realized with six markers. This combination presented the best discriminatory power (i.e

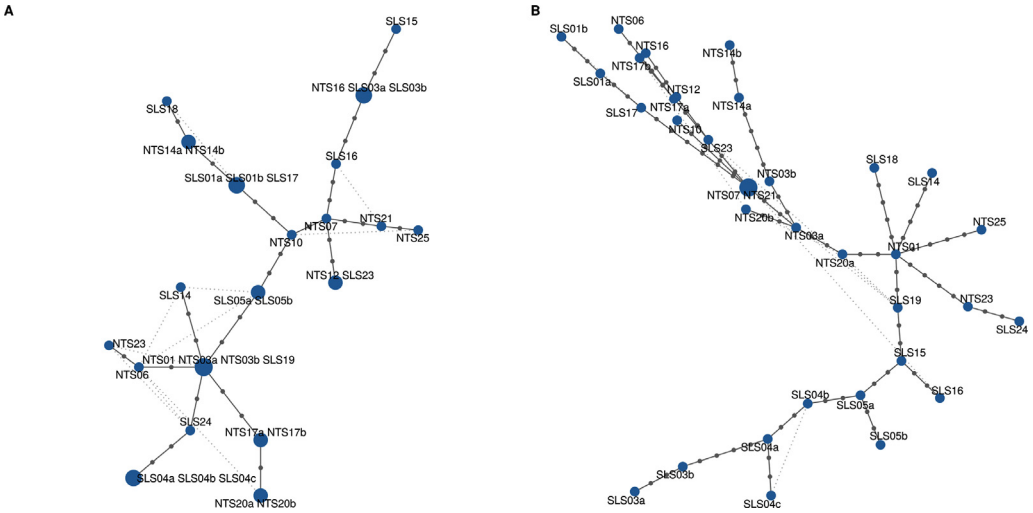

Fig. 2. A: MLST Haplotype network inferred with concatenated sequences data of the nucleotide polymorphic sites at the four loci (*ITS1, mt26S, SOD and CYB*). Fig. 2 B: MLP Haplotype network inferred with amplicon length data of the six microsatellites markers (STR022, STR108, STR138, STR189, STR278, and STR279). Blue circle = Haplotype/Genotype described within the sample (name of the sample, followed by a letter if several genotypes described in this sample). The dots separating two haplotypes correspond to the number of mutational steps needed to relate two distinct haplotypes. SLS: Saint-Louis; NTS: Nantes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.995 in this study and 0.992 in the European study[5]). Using the six markers is therefore recommended, acknowledging that analyzing six markers, using the same PCR protocol, is technically easily feasible, with a reasonable cost.

Five supplementary genotypes were described with MLP. They were from samples harboring mixed genotypes at only one STR locus. Indeed, one of the main findings of this study is a higher potential of MLP as compared with MLST for detecting mixtures of distinct genotypes in clinical samples. This mixture detection has a direct impact on typeability performance and MLST present a better typeability than MLP (86.5% versus 78.3%). However not detecting mixtures is preventing the detection of minority alleles associated with a minority genotype present in the specimen of the patient who could be part of the transmission chain of a specific genotype [10]. A better mixture detection was expected considering that 70% of PCP were mixtures as detected with our MLP method and that>90% of PCP are due to mixtures based on NGS analysis [8]. Indeed, MLST cannot exclude the presence of minor genotypes at a minimal ratio of 1:3 in a mixture due to technical issues of the Sanger sequencing [31], whereas MLP can detect minority (>2%) genotypes at a minimal ratio of 1:50 [12].

Only alleles from marker STR279 exhibit a widespread distribution, with the 178 bp allele linked to Nantes samples and the 181 bp allele linked to Paris samples. This result, observed for only one marker have already been reported in our European multicenter study (n = 249). Indeed, STR279 repartition appeared to be the most different between centers ($p < 0.001$), but differences were also observed for STR278 ($p = 0.006$) and STR138 ($p = 0.011$) [5]. Moreover, we could not exclude differences in rate of genetic modification events across microsatellites markers, with regards to their genome localization, or their nucleotide motif composition in A/T or G/C and their length (di-/tri- or tetra-nucleotide motif) [16,32–34]. The fact that we observed already known genotypes, even from others countries is consistent with previous findings on *P. jirovecii* population structure, i.e. a huge variety of genotypes but with a limited global population structure , as described also by Parobek and colleagues in their study with samples from Ouganda, Spain and USA [5,35].

Given the high number of mutational steps separating the different haplotypes, networks may be characterized by low support, since alternative branching may include a number of unobserved haplotypes in this dataset [24]. To this respect, the MLP genotyping appears more resolutive considering the higher number of mutational steps necessary between nodes (i.e. a total of 81 for MLP vs. 30 for MLST).

From the practical point of view, the theoretical reproducibility of MLST is well established and allows sequence exchanges among centers and available databases to compare genotypes. By comparison, when using MLP, it is recommended to use a standard allelic ladder to ensure good inter-laboratory reproducibility [16,36]. In addition, no database for MLP has been implemented yet.

Major criteria that should be considered to choose a genotyping method, such as discriminatory power, theorical reproducibility, feasibility, cost and time results are fulfilled by both methods, which can be used, depending on local facilities to study outbreaks. MLP is cheaper, and less time-consuming, with only one step of amplification and no need of sequencing. However, MLST could be easier to determine genotype with less subjective interpretation of amplicon size [37].

Depending on the aim behind the implementation of a genotyping study (outbreak investigation versus population structure study), one method could be prefered over another. If the aim is to investigate a local outbreak and search for interhuman transmission of a specific clone, the MLST or MLP approaches are convenient. However, if the aim is to study the physiopathology, the variability of infective strains or *P. jirovecii* population diversity, MLP will be the best method. The choice to use mitochondrial markers in MLST scheme could add some polymorphism and increase diversity index but these polymorphisms should not be linked to nuclear polymorphism [8].

In conclusion, both methods can be used, but MLST may not fit the biological diversity of the genome using the current scheme. Based on this observation, we could hypothesize that further developments in next generation sequencing technologies in a close future, restricted to nuclear genome will probably change the way we perform genotyping for this pathogen.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

MGM's salary was funded by Assistance Publique-Hôpitaux de Paris and Institut Pasteur (Poste d' accueil APHP/CNRS/Institut Pasteur). http://recherche-innovation.aphp.fr/2019/07/11/resultats-de-lappel-a-candidatures-poste-daccueil-2019-de-lap-hp/.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.10.005.

References

[1] Cordonnier C, Cesaro S, Maschmeyer G, Einsele H, Donnelly JP, Alanio A, et al. Pneumocystis jirovecii pneumonia: still a concern in patients with haematological malignancies and stem cell transplant recipients. J Antimicrob Chemother 2016;71:2379–85. https://doi.org/10.1093/jac/dkw155.

[2] Bitar D, Lortholary O, Le Strat Y, Nicolau J, Coignard B, Tattevin P, Che D, Dromer F. Population-Based Analysis of Invasive Fungal Infections, France, 2001–2010. Emerg Infect Dis 2014;20(7):1163–9. https://doi.org/10.3201/eid2007.140087.

[3] Alanio A, Desnos-Ollivier M, Garcia-Hermoso D, Bretagne S. Investigating Clinical Issues by Genotyping of Medically Important Fungi: Why and How?. Clin Microbiol Reviews 2017;30(3):671–707. https://doi.org/10.1128/CMR.00043-16.

[4] de Boer MGJ, de Fijter JW, Kroon FP. Outbreaks and clustering of *Pneumocystis* pneumonia in kidney transplant recipients: a systematic review. Med Mycol 2011:1–8. https://doi.org/10.3109/13693786.2011.571294.

[5] Alanio A, Gits-Muselli M, Guigue N, Desnos-Ollivier M, Calderon EJ, Di Cave D, Dupont D, Hamprecht A, Hauser PM, Helweg-Larsen J, Kicia M, Lagrou K, Lengerova M, Matos O, Melchers WJG, Morio F, Nevez G, Totet A, White LP, Bretagne S. Diversity of Pneumocystis jirovecii Across Europe: A Multicentre Observational Study. EBioMedicine 2017;22:155–63. https://doi.org/10.1016/j.ebiom.2017.06.027.

[6] Hauser PM, Blanc DS, Sudre P, Manoloff ES, Nahimana A, Bille J, et al. Genetic diversity of Pneumocystis carinii in HIV-positive and -negative patients as revealed by PCR–SSCP typing: AIDS 2001;15:461–6. Doi: 10.1097/00002030-200103090-00004.

[7] Esteves F, Gaspar J, De Sousa B, Antunes F, Mansinho K, Matos O. Clinical Relevance of Multiple Single-Nucleotide Polymorphisms in Pneumocystis jirovecii Pneumonia: Development of a Multiplex PCR-Single-Base-Extension Methodology. J Clin Microbiol 2011;49(5):1810–5. https://doi.org/10.1128/JCM.02303-10.

[8] 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. Doi: 10.3389/fmicb.2016.00733.

[9] Esteves F, Montes-Cano MA, Horra Cdl, Costa MC, Calderón EJ, Antunes F, Matos O. Pneumocystis jirovecii multilocus genotyping profiles in patients from Portugal and Spain. Clin Microbiol Infect 2008;14(4):356–62. https://doi.org/10.1111/j.1469-0691.2007.01944.x.

[10] Gits-Muselli M, Peraldi M-N, de Castro N, Delcey V, Menotti J, Guigue N, 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. Doi: 10.1371/journal.pone.0125763

[11] 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(9):2843–9. https://doi.org/10.1128/JCM.01073-13.
[12] Robin C, Alanio A, Gits-Muselli M, la Martire G, Schlemmer F, Botterel F, et al. Molecular Demonstration of a Pneumocystis Outbreak in Stem Cell Transplant Patients: Evidence for Transmission in the Daycare Center. Front Microbiol 2017;8. Doi: 10.3389/fmicb.2017.00700.
[13] Garcia-Hermoso D, Cabaret O, Lecellier G, Desnos-Ollivier M, Hoinard D, Raoux D, Costa J-M, Dromer F, Bretagne S. Comparison of Microsatellite Length Polymorphism and Multilocus Sequence Typing for DNA-Based Typing of Candida albicans. J Clin Microbiol 2007;45(12):3958–63. https://doi.org/10.1128/JCM.01261-07.
[14] Vanhee LME, Symoens F, Jacobsen MD, Nelis HJ, Coenye T. Comparison of multiple typing methods for Aspergillus fumigatus. Clin Microbiol Infect 2009;15(7):643–50. https://doi.org/10.1111/j.1469-0691.2009.02844.x.
[15] Garcia-Hermoso D, Desnos-Ollivier M, Bretagne S. Typing Candida Species Using Microsatellite Length Polymorphism and Multilocus Sequence Typing. Methods Mol Biol Clifton NJ 2016;1356:199–214. https://doi.org/10.1007/978-1-4939-3052-4_15.
[16] de Valk HA, Meis JFGM, Klaassen CHW. Microsatellite based typing of Aspergillus fumigatus: Strengths, pitfalls and solutions. J Microbiol Methods 2007;69(2):268–72. https://doi.org/10.1016/j.mimet.2007.01.009.
[17] Alanio A, Desoubeaux G, Sarfati C, Hamane S, Bergeron A, Azoulay E, Molina JM, 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(10):1531–7. https://doi.org/10.1111/j.1469-0691.2010.03400.x.
[18] Meliani L, Develoux M, Marteau-Miltgen M, Magne D, Barbu V, Poirot J-L, et al. Real Time Quantitative PCR Assay for Pneumocystis jirovecii Detection. J Eukaryot Microbiol 2003;50:651–651. Doi: 10.1111/j.1550-7408.2003.tb00670.x
[19] Tsolaki AG, Miller RF, Underwood AP, Banerji S, Wakefield AE. 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(1):141–56. https://doi.org/10.1093/infdis/174.1.141.
[20] Hauser PM, Francioli P, Bille J, Telenti A, Blanc DS. Typing of Pneumocystis carinii f. sp. hominis by single-strand conformation polymorphism of four genomic regions. J Clin Microbiol 1997;35:3086–91
[21] van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 2007;13:1–46. https://doi.org/10.1111/j.1469-0691.2007.01786.x.
[22] Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 1992;132:619–33.
[23] Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinforma Oxf Engl 2019;35:526–8. Doi: 10.1093/bioinformatics/bty633.
[24] Paradis E, Pearse W. Analysis of haplotype networks: The randomized minimum spanning tree method. Methods Ecol Evol 2018;9(5):1308–17. https://doi.org/10.1111/2041-210X.12969.
[25] Severiano A, Carriço JA, Robinson DA, Ramirez M, Pinto FR. Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PloS One 2011;6:e19539. Doi: 10.1371/journal.pone.0019539.
[26] Alanio A, Olivi M, Cabaret O, Foulet F, Bellanger A-P, Millon L, Berceanu A, Cordonnier C, Costa J-M, Bretagne S. Correlation Between Pneumocystis jirovecii Mitochondrial Genotypes and High and Low Fungal Loads Assessed by Single Nucleotide Primer Extension Assay and Quantitative Real-Time PCR. J Eukaryot Microbiol 2015;62(5):650–6. https://doi.org/10.1111/jeu.12222.
[27] Valero C, Buitrago MJ, Gits-Muselli M, Benazra M, Sturny-Leclère A, Hamane S, et al. Copy Number Variation of Mitochondrial DNA Genes in Pneumocystis jirovecii According to the Fungal Load in BAL Specimens. Front Microbiol 2016;7. Doi: 10.3389/fmicb.2016.01413.
[28] Montooth KL, Rand DM. The Spectrum of Mitochondrial Mutation Differs across Species. PLoS Biol 2008;6:e213. Doi: 10.1371/journal.pbio.0060213.
[29] Basse CW. Mitochondrial inheritance in fungi. Curr Opin Microbiol 2010;13(6):712–9. https://doi.org/10.1016/j.mib.2010.09.003.
[30] Westermann B. Mitochondrial inheritance in yeast. Biochim Biophys Acta (BBA) - Bioenerget 2014;1837(7):1039–46. https://doi.org/10.1016/j.bbabio.2013.10.005.
[31] Angulo B, García-García E, Martínez R, Suárez-Gauthier A, Conde E, Hidalgo M, López-Ríos F. A Commercial Real-Time PCR Kit Provides Greater Sensitivity than Direct Sequencing to Detect KRAS Mutations. J Mol Diagn 2010;12(3):292–9. https://doi.org/10.2353/jmoldx.2010.090139.
[32] Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci 1997;94(3):1041–6. https://doi.org/10.1073/pnas.94.3.1041.
[33] Bachtrog D, Agis M, Imhof M, Schlötterer C. Microsatellite Variability Differs Between Dinucleotide Repeat Motifs—Evidence from Drosophila melanogaster. Mol Biol Evol 2000;17:1277–85. Doi: 10.1093/oxfordjournals.molbev.a026411
[34] Bell GI. Evolution of simple sequence repeats. Comput Chem (Oxford) 1996;20(1):41–8. https://doi.org/10.1016/s0097-8485(96)80006-4.
[35] Parobek CM, Jiang LY, Patel JC, Alvarez-Martinez MJ, Miro JM, Worodria W, Andama A, Fong S, Huang L, Meshnick SR, Taylor SM, Juliano JJ. Multilocus Microsatellite Genotyping Array for Investigation of Genetic Epidemiology of Pneumocystis jirovecii. J Clin Microbiol 2014;52(5):1391–9. https://doi.org/10.1128/JCM.02531-13.
[36] Garcia-Hermoso D, MacCallum DM, Lott TJ, Sampaio P, Serna M-J- B, Grenouillet F, Klaassen CHW, Bretagne S. Multicenter Collaborative Study for Standardization of Candida albicans Genotyping Using a Polymorphic Microsatellite Marker. J Clin Microbiol 2010;48(7):2578–81. https://doi.org/10.1128/JCM.00040-10.
[37] Pasqualotto AC, Denning DW, Anderson MJ. A Cautionary Tale: Lack of Consistency in Allele Sizes between Two Laboratories for a Published Multilocus Microsatellite Typing System. J Clin Microbiol 2007;45(2):522–8. https://doi.org/10.1128/JCM.02136-06.