Multicentre study highlighting clinical relevance of new high-throughput methodologies in molecular epidemiology of Pneumocystis jirovecii pneumonia

F. Esteves¹, B. de Sousa², E. J. Calderón³, L. Huang⁴, R. Badura⁵, F. Maltez⁶, Q. Bassat⁷, Y. de Armas⁹, F. Antunes¹⁰ and O. Matos¹¹

¹) NOVA Medical School/Faculdade de Ciências Médicas, Centro de Toxicogenómica e Saúde Humana (ToxOmics), Departamento de Genética, Universidade Nova de Lisboa, Lisbon, 2) Faculdade de Psicologia e de Ciências da Educação, Coimbra, Portugal, 3) CIBER de Epidemiologia y Salud Publica and Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocio/CSIC/Universidad de Sevilla, Seville, Spain, 4) Division of Pulmonary and Critical Care Medicine & HIV/AIDS Division, San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, CA, USA, 5) Centro Hospitalar de Lisboa Norte, Hospital de Santa Maria, Faculdade de Medicina, Universidade de Lisboa, 6) Centro Hospitalar de Lisboa Central, Hospital de Curry Cabral, Serviço de Doenças Infecciosas, Lisbon, Portugal, 7) ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clinic Universitat de Barcelona, Barcelona, Spain, 8) Centro de Investigación em Saúde de Maniça (CISM), Maputo, Mozambique, 9) Laboratory of Molecular Diagnostic, Pedro Kouri Hospital, Havana, Cuba, 10) Instituto de Saúde Ambiental, Faculdade de Medicina da Universidade de Lisboa and 11) Unidade de Parasitologia Médica, Grupo de Protozoários Oportunistas/VIH e Outros Protozoários, Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal

Abstract

Pneumocystis jirovecii causes severe interstitial pneumonia (PcP) in immunosuppressed patients. This multicentre study assessed the distribution frequencies of epidemiologically relevant genetic markers of P. jirovecii in different geographic populations from Portugal, the USA, Spain, Cuba and Mozambique, and the relationship between the molecular data and the geographical and clinical information, based on a multifactorial approach. The high-throughput typing strategy for P. jirovecii characterization consisted of DNA pooling using quantitative real-time PCR followed by multiplex-PCR/single base extension. The frequencies of relevant P. jirovecii single nucleotide polymorphisms (mt85, SOD110, SOD215, DHFR312, DHPS165 and DHPS171) encoded at four loci were estimated in ten DNA pooled samples representing a total of 182 individual samples. Putative multilocus genotypes of P. jirovecii were shown to be clustered due to geographic differences but were also dependent on clinical characteristics of the populations studied. The haplotype DHFR312T/SOD110C/SOD215T was associated with severe AIDS-related PcP and high P. jirovecii burdens. The frequencies of this genetic variant of P. jirovecii were significantly higher in patients with AIDS-related PcP from Portugal and the USA than in the colonized patients from Portugal, and Spain, and children infected with P. jirovecii from Cuba or Mozambique, highlighting the importance of this haplotype, apparently associated with the severity of the disease and specific clinical groups. Patients from the USA and Mozambique showed higher rates of DHPS mutants, which may suggest the circulation of P. jirovecii organisms potentially related with trimethoprim-sulfamethoxazole resistance in those geographical regions. This report assessed the worldwide distribution of P. jirovecii haplotypes and their epidemiological impact in distinct geographic and clinical populations.

© 2016 European Society of Clinical Microbiology and Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

Keywords: DNA pools, high-throughput molecular epidemiology, immunosuppressed patients, multilocus genotyping survey, Pneumocystis jirovecii, pneumonia

Original Submission: 29 December 2015; Revised Submission: 19 February 2016; Accepted: 13 March 2016

Editor: E. Roilides

Article published online: 26 March 2016

Corresponding author: O. Matos, Unidade de Parasitologia Médica, Grupo de Protozoários Oportunistas/VIH e Outros Protozoários, Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira N° 100, 1349-008 Lisbon, Portugal
E-mail: omatos@ihmt.unl.pt
Introduction

Pneumocystis jirovecii pneumonia (PCP) is a major concern among human immunodeficiency virus (HIV)-infected persons and non-HIV-infected persons who are undergoing immunosuppressive treatments related to malignancies, connective tissue diseases or organ transplantation [1–3]. Pulmonary colonization with P. jirovecii in patients presenting with diverse levels of immunodeficiency, primary respiratory disorders, or even in the immunocompetent general population, is also an important epidemiological issue, especially in terms of transmission [4–9].

Although a culture system to propagate P. jirovecii in vitro was developed in 2014, it still needs to be validated, disseminated and shown to be cost-effective for diagnostic purposes [10]. In the absence of a well-established culture system to isolate and maintain live organisms, previous efforts were made to understand the patterns of transmission so as to develop methods of detection, intervention, characterization and management for P. jirovecii [11–19]. Recently, the de novo assembly of the P. jirovecii genome was published, opening the way to solve some critical issues, such as the identification of nutritional supplements for the development of reliable and cost-effective culture in in vitro systems, and detection of new targets for development of anti-PCP drugs and vaccines [20]. The multiplex amplification of genomic P. jirovecii DNA associated with single base extension (SBE) and DNA pooling was reported to be a reliable alternative high-throughput DNA sequencing technique, allowing the calculation of the single nucleotide polymorphism (SNP) allele frequencies in a large number of samples [21,22]. DNA pooling is a reliable and time-saving method for genotyping screenings, in which equal amounts of DNA from a large number of individual samples are pooled and the SNP allele frequencies are estimated [21].

In the last two decades, several P. jirovecii DNA regions were studied and specific SNPs associated with parameters of P. jirovecii infection were identified [2,3,14–18,23–25]. The mitochondrial large subunit ribosomal RNA (mtLSUrRNA) is a conserved multicopy gene with a central role in basic metabolic mechanisms during translation, providing peptidyl transferase activity to the mitochondrial ribosome [11,14,18,19,22]. Genetic variations at base 85 of the mtLSUrRNA gene were recognized to be potentially associated with high P. jirovecii burden levels and unfavourable follow up of infection [21,22]. The dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) genes encode for two P. jirovecii central enzymes in the folate synthesis. Significant associations between the SNP at bases 165 and 171 of DHPS and the use of sulfa drugs for PCP prophylaxis [2,15,17,25] or failure of both trimethoprim-sulfamethoxazole treatment [26] and trimethoprim-sulfamethoxazole or dapsone prophylaxis [3], were reported. Additionally, P. jirovecii may evolve under pressure from DHFR inhibitors, such as trimethoprim or pyrimethamine, and mutations in this gene may contribute to drug resistance [23,24,27]. The SNP at position 312 was associated previously with PCP infection burden [18]. Superoxide dismutase (SOD) is involved in the protective mechanisms of P. jirovecii against reactive oxygen radicals produced by alveolar macrophages or neutrophils [22,28]. The major genetic variations of the SOD locus (bases 110 and 215) are reported to be at linkage disequilibrium and associated with severity of PCP episodes [18,21,22,28].

This report is the first multicentre P. jirovecii molecular study in different geographic populations from four different continents. The aims were: to evaluate the distribution frequencies of specific genetic markers in four P. jirovecii loci, in populations from five different geographic origins (Portugal, the USA, Spain, Cuba and Mozambique), including important genomic regions involved in basic metabolic mechanisms, such as the mtLSUrRNA, SOD, DHFR and DHPS and to epidemiologically assess the relationship between the molecular data and the geographical and clinical information.

Materials and Methods

Subjects and data

A cohort of 182 respiratory specimens tested previously and found to be positive for P. jirovecii by real-time quantitative PCR (qPCR) were included in the study. Specimens included bronchoalveolar lavage fluids and induced sputa (adult patients) or nasopharyngeal swabs (children) collected during routine diagnostic procedures/clinical care in five different geographical locations (multicentre study). In each healthcare/diagnostic centre, data were collected using standardized data collection forms. The present study had the approval of the Institutional Review Boards/Ethical Committees from the involved institutions. The clinical and demographic data are summarized in Table 1.

Pneumocystis jirovecii burden was quantified in the 70 AIDS-related PCP episodes from Portugal by scoring the number of cysts observed by applying the semi-quantitative method of indirect immunofluorescence staining with monoclonal antibodies (MonoFluoTM kit P. jirovecii; Bio-Rad, Marnes-la-Coquette, France) and designated as low/moderate (one to three cysts in one field at × 1000) in 38 cases, and as heavy (four or more cysts in one field at × 1000) in 32 cases. Follow up was possible in 53 of the 90 Portuguese patients with PCP. A follow up was considered positive when the patient showed a
TABLE 1. Clinical information and demographic data of the groups of patients involved in the study

<table>
<thead>
<tr>
<th>Country (area)</th>
<th>Patients</th>
<th>Mean age (range) years</th>
<th>Male to female sex ratio</th>
<th>Sampling collection, month year</th>
<th>Type of respiratory sample</th>
<th>Immune status</th>
<th>Pneumocystis jirovecii detection</th>
<th>Clinical data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal (Lisbon)</td>
<td>108</td>
<td>40 (18–79)</td>
<td>1.78:1</td>
<td>March 2004 February 2012</td>
<td>88 BAL 20 IS</td>
<td>88 HIV (+) 20 HIV (−) 1</td>
<td>Microscopy (IF/mAb) and nPCR</td>
<td>90 microscopically confirmed PcP (70 HIV (+), 20 HIV (−)); 18 microscopically-negative for PcP, positive by nPCR (other pulmonary diseases)</td>
</tr>
<tr>
<td>USA (San Francisco, CA)</td>
<td>30</td>
<td>Unavailable 1</td>
<td>Unavailable 1</td>
<td>February 2004 December 2012</td>
<td>13 BAL 17 IS</td>
<td>30 HIV (+)</td>
<td>Microscopy (modified Giemsa stain, Diff-Quik) and nPCR</td>
<td>Microscopically confirmed PcP receiving mechanical ventilation and accompanying sedation</td>
</tr>
<tr>
<td>Mozambique (Maputo)</td>
<td>22</td>
<td>3 months (1–3 months)</td>
<td>1.75:1</td>
<td>November 2006 October 2007</td>
<td>22 NP swabs</td>
<td>9 HIV (+) 6 HIV (−) severely malnourished 7 Unknown 0</td>
<td>Microscopy (Giemsa stain) and nPCR</td>
<td>Microscopically negative for PcP, positive by nPCR presenting cough, fever and distress</td>
</tr>
<tr>
<td>Spain (Seville)</td>
<td>12</td>
<td>51 (18–87)</td>
<td>3:1</td>
<td>January 2006 July 2013</td>
<td>7 BAL 5 IS</td>
<td>7 HIV (+) 5 HIV (−) COPD</td>
<td>Microscopy (Giemsa stain) and nPCR</td>
<td>Microscopically negative for PcP, positive by nPCR colonized by P. jiroveci and diagnosed with other pulmonary diseases than PcP</td>
</tr>
<tr>
<td>Cuba (Havana)</td>
<td>10</td>
<td>5 months (1–10 months)</td>
<td>0.25:1</td>
<td>July 2013 August 2013</td>
<td>10 NP swabs</td>
<td>10 HIV (−)</td>
<td>Microscopy (Giemsa stain) and nPCR</td>
<td>Microscopically-negative for PcP, positive by nPCR presenting cough, fever and respiratory secretions</td>
</tr>
</tbody>
</table>

Abbreviations: HIV (+), HIV-positive patients; HIV (−), HIV-negative patients; BAL, bronchoalveolar lavage; IS, induced sputum; IF/mAb, indirect immunofluorescence staining with monoclonal antibodies; nPCR, nested PCR directed to P. jiroveci mtLSU rRNA gene; NP, nasopharyngeal. COPD, chronic obstructive pulmonary disease.

1Six patients with neoplasia, five organ transplantation recipients and nine patients with no available data to establish immune status.

2 Patients with no available data to establish immune status.

3 Young children infected with P. jiroveci, presenting respiratory symptoms. Due to the low sensitivity of the NP swab, detection of P. jiroveci was possible only by nPCR.

favourable response to anti-P. jiroveci therapy and survived for at least 4 weeks after the diagnosis of PcP. Negative follow up was established either when there was a negative response to anti-P. jiroveci therapy (failure to improve clinically after administration of the drug for more than 10 days) or when the patient died during a PcP episode [18,21].

**DNA pooling**

After collection, the samples from the five cohorts were processed and immediately stored at −20°C for further analysis. All specimens were subjected to DNA extraction using the Qiap kit (Qiagen, Hilden, Germany). Molecular detection of P. jiroveci was performed by nested-PCR (nPCR) directed to the P. jiroveci large subunit mitochondrial rRNA (mtLSU rRNA) gene [7,12,29]. All respiratory specimens were confirmed to be positive for P. jiroveci by qPCR targeting the kexin-like serine protease (KEX1) gene of P. jiroveci [13,21].

**Pneumocystis jiroveci** DNA pools were planned based on geographical origin and clinical data. DNA quantification of each individual respiratory specimen was achieved using the qPCR targeting the KEX1 gene of P. jiroveci. The assay was performed in the 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), using the TaqMan® Gene Expression with minor groove binder probes FAM™ dye-labelled (Applied Biosystems): 2 min at 50°C, 10 min at 95°C and 50 amplification cycles of 15 s at 95°C and 1 min at 60°C using a 9-μL DNA sample, 1× TaqMan® Gene Expression Master Mix (Applied Biosystems), 1× TaqMan® Gene Expression Assay (forward primer 5′-CAACCCCTGTTCATCGCCTA-3′, reverse primer 5′-CAA-CACCGATTCACAAAACAGT-3′ and minor groove binder probe 5′-TGCTGTTAGCTTGGCCCTA-3′; Applied Biosystems), in a 20-μL reaction volume. The baseline was taken from cycles three to 15 and the threshold was set at 0.02. The amount of P. jiroveci DNA present in each individual sample was calculated applying the standard P. jiroveci DNA pattern curve $C_T = -3.4323 \log_{10} [KEX1] + 20.3610$, in which $C_T$ is the quantification cycles and [KEX1] is the concentration of the KEX1 fragments (ng/mL). This standard curve represents the relationship between P. jiroveci KEX1 gene fragment concentration and qPCR $C_T$ values, previously estimated using serial dilutions of KEX1 PCR product suspensions quantified by the PicoGreen dsDNA quantification reagent method [13,21].

The respiratory specimens were diluted (1: 20) and the respective $C_T$ values were estimated and converted into P. jiroveci DNA concentration (ng/mL) using the standard curve. The concentration of KEX1 gene copies (copies/μL) was derived from the
PicoGreen dsDNA quantification reagent method and KEX1 DNA fragment molecular weight (229.091.67 g/mol). As the KEX1 is a nuclear single-copy gene, the number of copies per microtitre corresponds to the P. jirovecii genome concentration. The average genome concentrations (P. jirovecii genomes/μL) of the pooled P. jirovecii samples were PT1 2.89 × 10⁶, PT2 8.74 × 10⁶, PT3 2.00 × 10⁷, PT4 1.69 × 10⁷, PT5 7.59 × 10⁷, PT6 2.51 × 10⁷, USA 1.42 × 10⁷, MOZ 1.74 × 10⁷, SPA 1.85 × 10⁷ and CUB 1.13 × 10⁸. Equivalent amounts of DNA (1 × 10⁻⁵ ng) from each of the individual samples were proportionally combined in the corresponding pool, according to geographical origin and/or clinical data, as follows.

**Portugal Pool 1 (PT1).** Thirty-eight respiratory specimens from HIV-positive adult patients with AIDS-related PcP (positive microscopy, positive nPCR), presenting low/moderate parasite burden.

**Portugal Pool 2 (PT2).** Thirty-two respiratory specimens from HIV-positive adult patients with AIDS-related PcP (positive microscopy, positive nPCR), presenting high parasite burden.

**Portugal Pool 3 (PT3).** Eighteen respiratory specimens from HIV-positive adult patients PcP-negative (negative microscopy, positive nPCR) with other pulmonary diseases, colonized by P. jirovecii (subclinical infection).

**Portugal Pool 4 (PT4).** Twenty respiratory specimens from HIV-negative adult patients with PcP (positive microscopy, positive nPCR).

**Portugal Pool 5 (PT5).** Thirty-five respiratory specimens from HIV-positive adult patients with PcP, presenting clinical improvement (positive follow up).

**Portugal Pool 6 (PT6).** Eighteen respiratory specimens from HIV-positive adult patients with PcP who failed to improve clinically or died during the PcP episode (negative follow up).

**USA Pool.** Thirty respiratory specimens from HIV-positive patients with severe AIDS-related PcP (positive microscopy, positive nPCR).

**Mozambique Pool.** Twenty-two respiratory specimens from infants infected with P. jirovecii, presenting respiratory symptoms (negative microscopy, positive nPCR).

**Spain Pool.** Twelve respiratory specimens from HIV-positive and HIV-negative adult patients colonized by P. jirovecii, presenting with other respiratory diseases than PcP (negative microscopy, positive nPCR).

**Cuba Pool.** Ten respiratory specimens from HIV-negative young children infected with P. jirovecii, presenting respiratory symptoms (negative microscopy, positive nPCR).

**Genotyping**

The P. jirovecii DNA pools were studied using the multiplex-PCR (MPCR)/SBE technique (in triplicate), as described previously [21,22]. Four P. jirovecii hot spots (mtLSUrRNA, SOD, DHFR and DHPS) were simultaneously amplified using an MPCR (T1 Thermocycler; Biometra, Göttingen, Germany), as follows: 10 min at 95°C, followed by 45 amplification cycles of 1 min at 95°C, 1 min 60°C and 1 min at 72°C, and a final extension of 10 min at 72°C, using 4 μL DNA sample, 2.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems), 1.5 × reaction buffer (75 mm KCl, 15 mm Tris–HCl (pH 8.3); Applied Biosystems), 0.5 mM deoxyribonucleoside triphosphates (dNTPs) (Applied Biosystems), 4 mM MgCl₂ (Applied Biosystems), 0.01 μg/μL bovine serum albumin (Sigma–Aldrich, Cleveland, OH, USA), 0.75 μL dimethylsulphoxide (DMSO) (Sigma–Aldrich), 0.5 μM of mtLSUrRNA primers (pAZ102-X and pAZ102-E), 1.4 μM of SOD primers (MnSODFw and MnSODRw2), 0.7 μM of DHPS primers (DHPSFw1 and DHPSRv1), and 1.4 μM of DHFR primers (FR 208 and FR 1018), in a 50-μL reaction volume. Except for DHPSFw1 (5′-GCATGGGGGTTCTATTCA-TATG-3′) and DHPSRv1 (5′-GCTTTTATTGGTTTCTTG-CAACC-3′), all primers were described previously [11,16,19,21–23,29].

The MPCR products (10 μL) were incubated with shrimp alkaline phosphatase (2 U) (USB Corporation, Cleveland, OH, USA) and exonuclease I (4 U) (USB Corporation) for 1 h at 37°C (20 μL reaction volume). After inactivation of the enzymes (15 min at 96°C), 5 μL of treated MPCR products were used in the SBE reaction (15 μL reaction volume): 1 min at 90°C, followed by 45 SBE cycles of 10 sec at 90°C and 20 sec 45°C, using 4 μL SNPStart Master Mix (GenomeLab SNPStart primer extension kit; Beckman Coulter, Brea, CA, USA) and SBE-TAG probes (2.7 μM SOD215, 2.1 μM SOD110, 0.9 μM DHPS171) (Eurofins Genomics, Ebersberg, Germany). Except for DHPS165 (5′-GGATAAATTCTACACACCGTGGTGTTGACTATTTTG-CCAAAGT-0, 5′-CAGAGCCATCTATAGAAACAACATSTGAACCAG-3′), in which 5′ corresponds to Deoxyoxinosine, all SBE-TAG probes were described previously [22]. The SBE products were treated with shrimp alkaline phosphatase (0.5 U) for 1 h at 37°C, followed by enzyme inactivation (15 min at 96°C). The MPCR/SBE products were analysed in a CEQ 8000-XL (Beckman Coulter).
D3-green guanine and 50 nt D4-blue thymine) was run in each SBE assay. The average normalized relative frequencies of the SNP alleles in each DNA pool sample were calculated by dividing the maximum height values of fluorescence peaks observed in the SBE products by the reference fluorescence values of the positive control.

Data analysis
To overcome the failure of normality necessary for the application of Student’s t-test, the Kruskal–Wallis non-parametric test was used to analyse the differences in the SNP frequency distribution variation across the pools. The significance level considered in all the statistical tests was 0.05 [21,22].

The most representative multilocus genotypes (MLG) (SNP frequencies >32% for mt85 and 50% for the remaining SNP) of each pool were analysed and a dendrogram was computed using the software CLUSTAL W2 multiple sequence alignment (version 2.0.12).

Results

DNA pooling
According to the Kruskal–Wallis test, the median Cq values and genomes concentration were statistically different between several pools: PT1 versus PT2 (p 0.011); PT2 versus PT3 (p <0.001); PT2 versus PT4 (p <0.001); PT3 versus PT5 (p 0.002); PT3 versus USA (p 0.003); PT4 versus MOZ (p <0.001); PT4 versus SPA (p 0.035); PT4 versus CUB (p 0.011).

Genotyping
In all, 182 pulmonary specimens divided into ten DNA pools were analysed for the epidemiological distribution of P. jirovecii genotypes in distinct geographic regions and among patients presenting with different clinical conditions. The six SNP, located in the four genetic loci studied, were successfully characterized by the Multiplex-SBE/DNA pooling method in all DNA pools. Fluorescence peaks with 35 nt (DHFR312), 47 nt (mt85), 55 nt (DHPS165), 64 nt (SOD215), 75 nt (SOD110) and 82 nt (DHPS171) were detected in the SBE reactions and the average normalized relative frequencies of each SNP allele were estimated (Table 2, Fig. 1). Data analysis demonstrated several significant statistical differences in the frequency distribution of SNP among pools from different geographic origins and with distinct parameters of infection (Table 3).

The relationships between the most relevant putative MLG identified in the pools studied were analysed applying the neighbour joining method. The most representative putative MLG (SNP frequencies >32% for mt85 and 50% for the

| Table 2. Pneumocystis jirovecii single nucleotide polymorphism alleles relative frequencies distribution among the different pools studied |
|---|---|---|---|---|---|---|---|
| Pools | mt85A | mt85T | DHPS165A | DHPS165G | mt85C | mt85G |
| PT1 | 40.1 | 59.9 | 92.1 | 7.9 | 91.4 | 8.6 |
| PT2 | 84.9 | 15.1 | 92.1 | 7.9 | 91.4 | 8.6 |
| PT3 | 74.3 | 25.7 | 93.9 | 6.1 | 91.1 | 8.9 |
| PT4 | 79.3 | 20.7 | 84.8 | 15.2 | 96.7 | 3.3 |
| USA | 62.7 | 37.3 | 80.5 | 19.5 | 92.6 | 7.4 |
| MOZ | 63.7 | 36.3 | 82.1 | 17.9 | 82.8 | 17.2 |
| SPA | 66.6 | 33.4 | 82.1 | 17.9 | 82.8 | 17.2 |
| CUB | 60.0 | 40.0 | 82.1 | 17.9 | 82.8 | 17.2 |

Abbreviations: PcP, Pneumocystis jirovecii pneumonia; PT1-6, AIDS-related PcP Portuguese patients with different parasite burden; PT1, low parasite burden; PT2, high parasite burden; PT3, positive follow-up positive control; PT4, positive follow-up positive control; USA, AIDS-related PcP patients from USA; MOZ, Mozambican infants infected with P. jirovecii presenting respiratory symptoms; SPA, HIV-negative older children colonized by P. jirovecii presenting other respiratory symptoms than PcP; CUB, HIV-negative young children infected with P. jirovecii presenting respiratory symptoms. The relative frequency of each polymorphism was calculated by dividing the maximum height values of fluorescence peaks observed in the single-base extension products by the reference fluorescence observed in the positive control.
FIG. 1. Graphic representation of *Pneumocystis jirovecii* single nucleotide polymorphism alleles relative frequencies among the different pools studied. Radar charts were drawn with axes for each single nucleotide polymorphism (represented from the centre to the periphery) and a scale in the range of 0-100% (10% intervals).
remaining SNP) were considered in the construction of a specific dendrogram (Fig. 2) in which three major clusters were identified: cluster A with MLG from Spain (SPA-A, SPA-B), Cuba and Portugal (PT1-A, PT3); cluster B with MLG from Portugal (PT1-B, PT2, PT4, PT5, PT6); and cluster C with MLG from Mozambique and USA (USA-A, B, C, D).

**Table 3.** Multiple comparisons of the *Pneumocystis jirovecii* single nucleotide polymorphisms frequency distribution variation across the pools studied

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Multiple comparison analysis (p values)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>mt85A</td>
<td>PT1 vs. PT2 (0.0004); PT2 vs. SPA (0.001); SPA vs. MOZ (0.0005); PT1 vs. PT6 (0.036); PT2 vs. SPA (0.045); PT3 vs. SPA (0.013); PT2 vs. USA (0.019); PT1 vs. USA (0.018); PT3 vs. USA (0.002); USA vs. SPA (0.0008)</td>
</tr>
<tr>
<td>mt85C</td>
<td>PT1 vs. PT2 (0.0442); PT2 vs. USA (0.019); PT1 vs. USA (0.018); PT3 vs. USA (0.002); USA vs. SPA (0.0005)</td>
</tr>
<tr>
<td>DHFR117CT</td>
<td>PT2 vs. PT3 (0.012); PT3 vs. USA (0.015); USA vs. CUB (0.0009)</td>
</tr>
<tr>
<td>SOD1/107C</td>
<td>USA vs. CUB (0.0009)</td>
</tr>
<tr>
<td>SOD2/135T</td>
<td>PT2 vs. CUB (0.0032); USA vs. CUB (0.0009)</td>
</tr>
</tbody>
</table>

Abbreviations: PcP, *Pneumocystis jirovecii* pneumonia; PT1, AIDS-related PcP Portuguese patients with high parasite burden; PT2, AIDS-related PcP Portuguese patients with low/moderate parasite burden; PT3, HIV-positive PcP-negative Portuguese patients with other pulmonary diseases, colonized by *P. jirovecii*; PT4, HIV-negative Portuguese patients with PcP; PT5, Portuguese patients with PcP presenting positive follow-up; PT6, Portuguese patients with PcP presenting negative follow-up; SPA, AIDS-related PcP patients from Spain; MOZ, Mozambican infants infected with *P. jirovecii*, presenting respiratory symptoms; SPA, immunocompromised Spanish patients colonized by *P. jirovecii*, presenting other respiratory diseases than PcP; CUB, HIV-negative young children infected with *P. jirovecii*, presenting respiratory symptoms.

*For the multiple comparisons of the pools, the p values presented are the adjusted p values (Bonferroni correction) that took into consideration the number of comparisons, i.e. the adjusted value p = p value k/(k – 1/2), with k = 10 the number of pools to be compared is 45. The multiple comparisons were performed considering three groups of pools: (1) All ten pools; (2) only the six Portuguese pools; (3) only the four foreign pools. N.S, not significant; p > 0.05.

**Discussion**

To our knowledge, this is the first report on *P. jirovecii* mtLSU rRNA and DHFR gene variability in Mozambique, DHFR gene variability in Mozambique, Cuba and Spain, and SOD gene variability in the USA, Mozambique and Spain. The results strengthen the hypothesis that both geographic distances and clinical parameters have direct impact in the distribution of the genetic subtypes of *P. jirovecii* as demonstrated in Figs. 1 and 2.

In the Portuguese pools, data on genome concentrations are consistent with *P. jirovecii* burden, with the pool PT2 (AIDS-related PcP patients with high parasite burden) presenting the highest *P. jirovecii* genome concentration. The high concentration of *P. jirovecii* genomes detected in the pools from Cuba and Mozambique suggests that both populations were infected with high burdens of *P. jirovecii*, which was not initially detected, probably due to the low sensitivity and difficulty in reading of the microscopic diagnostic method using Giemsa staining [30].

**mtLSU rRNA**

In Portugal, mt85C has been reported as the most frequent allele of the SNP mt85 among HIV-positive patients with PcP (mt85C 58%–63.0%, mt85A 26.0%–43.0% and mt85T 18.0%–22.0%, time period 1997–2007) [12,18,19,22]. In the present study, except for the pool PT1, the Portuguese pools demonstrated mt85A as the most frequent allele. The allele mt85C was statistically associated with the pool PT1 (Tables 2 and 3). This finding is similar to previous reports, in which low/moderate *P. jirovecii* burden was more frequently observed among respiratory specimens with mt85C, whereas high burdens were more frequently detected in respiratory specimens with mt85A or mt85T [18,19,21]. The differences observed between the present and previous data may be the result of the different time periods of the studies (2004–2013 and 1997–2007, respectively).

In the USA, the distribution pattern of mt85 (mt85C 31.5%, mt85A 61.1% and mt85T 7.4%) was similar to previous studies (mt85C 42.9%–43.9%, mt85A 36.7%–50.0% and mt85T 7.1%–9.3%), in a population of PcP patients, time period 1986–1999 [14,31], particularly in San Francisco, with mt85C 25.0%, mt85A 50.0% and mt85T 7.0% (mixed genotypes 18.0%) (in HIV-positive patients) [14]. In the AIDS-related PcP patients from San Francisco, the distribution pattern of mt85 is relatively stable in different time periods (2004–2012 and 1995–1998).

Pools from Mozambique and Cuba showed similar distribution patterns of mt85 (mt85C 16.5%, mt85A 72.5%, mt85T 10.9% and mt85C 11.4%, mt85A 63.0%, mt85T 25.7%, respectively). Two recent Cuban studies, in a population of children colonized by *P. jirovecii* (both time periods 2010–2013) demonstrated analogous distribution patterns of mt85, with mt85C 18.0%–31.0%, mt85A 56.0%–68.0% and mt85T 25%–37% [5,6]. The similarity between the results from Mozambique and Cuba may be due to the fact that both pools were constituted by populations of young children infected with *P. jirovecii* and presenting analogous clinical parameters.

Since 2003, several studies in Spain demonstrated mtLSU rRNA variation patterns (mt85C 40.0%–55.1%, mt85A 10.1%–18.2% and mt85T 25.0%–36.7% in the general population, in HIV-positive patients with PcP, and in HIV-negative patients with other concomitant pulmonary disorders, between 1995 and 2008) [4,8,9,12]. Except for the general population study (mt85C and mt85A both most prevalent), mt85C was the most prevalent allele in all the other populations, followed by mt85T and mt85A. The present data showed a change in the allelic distribution of mt85 in Seville, Spain, particularly an increase of the mt85A. These differences appear to be more related to the periods in which the populations were studied (2006–2013 in the present study, and 1995–2008 in the earlier studies) than with the characteristics of the population, since the studies of Montes-
Cano et al. [4] and Esteves et al. [12] focused on Spanish populations similar to those in the present study.

**DHPS**

In Portugal, a decline of DHPS gene mutation frequencies has been documented since the beginning of the 2000s. Frequencies of 24.7% DHPS165G and 22.5% DHPS171T were detected in a population of HIV-positive and HIV-negative patients with pulmonary disorders, between 1994 and 2001, in which DHPS mutations were more frequent in the time period 1994–1997 (33%) than in 1998–2001 (9%) (p 0.022) [32]. More recently, the proportion of mutant DHPS alleles in a population of HIV-positive patients was 7.0% DHPS165G and 9.0% DHPS171T, between 2001 and 2007 [19]. The decline of DHPS mutant alleles among the HIV-positive patients (pools PT1, PT2 and PT3, time period 2004–2012) may be attributed to the lack of exposure to trimethoprim-sulfamethoxazole in this population, after the decreased use of sulfa prophylaxis due to widespread use of potent combination antiretroviral therapy in Europe, in the late 1990s [2,12,32]. However, in the HIV-negative patients with PCP (pool PT4), the frequencies of DHPS mutants, especially DHPS165G, appeared to be slightly higher than the ones reported in the HIV-positive patients. The reason for this observation is unclear. One possible explanation is that these groups are clinically distinct, reflecting different trimethoprim-sulfamethoxazole exposure, which may have an impact on the genetic variability of the DHPS locus in *P. jirovecii* [33,34].

In the USA, since 1998, several studies have reported DHPS mutations per PCP cases ranging from 26% to 81%, describing high overall frequencies of DHPS165G and DHPS171T in different time periods (7.4%–77.3% and 22.2%–59.4%, respectively, from 1976 to 2001 [3,14,15,17,31,35]), especially in San Francisco (65.0% DHPS165G, 64.0% DHPS171T) [18]. In this city, several reports consistently described overall high frequencies of DHPS mutations in the diseases gene (87.0%, 1995–1998 [14]; 81.5%, 1996–1999 [17]; 81.4%, 1997–2002 [15]). The high frequencies of DHPS mutants in the US pool (Fig. 1) is of concern and may reflect trimethoprim-sulfamethoxazole exposure of the AIDS-related PCP population and increased of sulfa-induced mutants, especially after the mid-1990s [3,14,23].

In the early 2000s, several studies described low frequencies of DHPS mutations in African countries neighbouring...
Mozambique (DHPS165G 7.1% and DHPS171T 0% in Zambian AIDS-related PCP patients, 1992–1993 [36]; DHPS165G 1.9% and 10.0%, DHPS171T 1.9% and 6.7% in South African adult patients with PCP and in HIV-infected children, respectively, 2000–2003 [37,38]). The low frequencies were attributed to the lack of exposure to trimethoprim-sulfamethoxazole in the populations studied [37,39]. However, in South Africa rates of DHPS165G 44.0% and DHPS171T 41.1% were reported in HIV-infected adult patients suspected of having PCP (time period 2006–2007) [38]. The results of the present study showed similar high prevalence of DHPS mutants in Mozambican infants with P. jirovecii infection in an overlapping time period (2006–2007). In this region of southeastern Africa, the difference in the percentage of P. jirovecii DHPS mutations may be due to different time periods of study, reflecting the increased widespread empirical use of trimethoprim-sulfamethoxazole, the mainstay of PCP treatment and prophylaxis regimens in sub-Saharan Africa, in the late 2000s [38]. But it can also be attributed to an increased awareness by clinicians leading to higher rates of diagnosis and subsequently higher detection rates of DHPS mutations.

In contrast to the USA and Mozambique, pools from Spain and Cuba showed low frequencies of DHPS mutants (Fig. 1). In Spain, the reported frequencies of DHPS mutants vary from 0% in the general population [8] to 15% for both DHPS165G and DHPS171T in HIV-positive and HIV-negative patients with pulmonary disorders [12]. One study involving HIV-negative patients with chronic pulmonary disease colonized by P. jirovecii showed frequencies of DHPS165G 21.4% and DHPS171T 14.3% (time period 2001–2002) [7]. Another study in AIDS-related PCP patients, detected frequencies of DHPS165G 16.1% and DHPS171T 12.8% (time period 2001–2003) [4]. These differences may be due to the lack of exposure to trimethoprim-sulfamethoxazole in the general population, when compared with AIDS-related PCP patients, who are more likely to be treated with that drug combination in prophylactic or therapeutic doses.

In Cuba, a recent study showed a frequency of 12.0% for both DHPS165G and DHPS171T in young children with whooping cough, colonized by P. jirovecii (between 2010 and 2013) [6]. This distribution pattern is consistent with the present results in Cuban HIV-negative young children infected with P. jirovecii, in a coincident time period (2013). The detection of low frequencies of DHPS mutations in Cuban children is most probably due to the lack of exposure of this specific population to trimethoprim-sulfamethoxazole.

**DHFR**

In Portugal, DHFR312T was the most prevalent DHFR allele, as has been found in previous studies [18,19,21]. The frequency of DHFR312C in AIDS-related PCP patients (PT1 and PT2) was similar to those previously found in Portuguese HIV-positive patients (21%–25%, 1997–2007 [19]; and 10%, 2001–2008 [21]). DHFR312C frequency was considerably higher in HIV-positive patients colonized by P. jirovecii (PT3) and HIV-negative PCP patients (PT4) that normally present lower P. jirovecii burdens than AIDS-related PCP patients (Fig. 1, Tables 2 and 3). The difference observed may be due to the recently reported association between DHFR312T and higher P. jirovecii burdens, also found in the present study (p 0.01) [18].

A very low frequency of DHFR polymorphic sequences was reported in the PCP patients from the USA, between 1985 and 1998 (2.7%) [23]. In the present study, the frequency of DHFR312C in the US patients was 14.8%. The reason for this increase is not obvious. It may eventually reflect the selective pressure of trimethoprim-induced polymorphisms caused by the increased widespread use of trimethoprim-sulfamethoxazole in the USA after the mid-1990s, or it may be linked to the association between DHFR312T and the higher P. jirovecii burdens normally found in AIDS-related PCP patients.

Similar distribution patterns of DHFR312 were observed between the pools from Mozambique, Spain and Cuba. In a retrospective study from South Africa, DHFR312C was not observed in a heterogeneous population of patients with pulmonary disorders (2000–2003) [39]. In Mozambique, the high levels of DHFR polymorphic sequences detected in the present study may reflect the increased widespread empirical use of trimethoprim-sulfamethoxazole in the late 2000s and the selective pressure of trimethoprim-induced polymorphisms.

**SOD**

In the present study, the distribution pattern of the SOD polymorphic sequences in the Portuguese population (SOD110C 8.8%–46.4% and SOD215T 8.0%–38.6%, time-period 2004–2012) was identical to a previous report on P. jirovecii multilocus genotyping in pooled DNA samples [21]. The highest frequencies of SOD110C (46.4%) and SOD215T (38.6%) were observed in the PT2 pool, supporting the hypothesis that the genotype SOD110C/SOD215T is linked to higher P. jirovecii burdens [18,21].

The US pool (AIDS-related PCP, most receiving mechanical ventilation) showed the most distinct distribution pattern of the SOD alleles (SOD110C 64.9%, SOD215T 50.6%). Again, the alleles SOD110C and SOD215T were statistically associated with this pool (Table 3), supporting also the relationship between the genotype SOD110C/SOD215T and more virulent PCP episodes.

The distribution patterns of the SOD alleles were similar among the pools from Mozambique (SOD110C 13.4%, SOD215T 12.6%), Spain (SOD110C 14.6%, SOD215T 8.6%) and Cuba (SOD110C 2.9%, SOD215T 2.4%). The pool from
Mozambique showed frequencies lower than those found in a previous study in Zimbabwean AIDS-related PcP patients, in 2004 (SOD110C 33.3%, SOD215T 33.3%) [28]. A possible explanation of this difference is related to the distinct clinical populations studied, reflecting the possible association between more virulent PcP episodes, reported in Zimbabwe, and the genotype SOD110C/SOD215T.

Low frequencies of SOD110C and SOD215T were observed in the pools from Spain and Cuba. However, in a recent study involving young Cuban children with whooping cough, colonized by P. jirovecii, the frequencies of SOD110C and SOD215T (time period 2010–2013) were much higher, 42.0% and 71%, respectively [6]. The reason for this discrepancy is not obvious. We hypothesize that this difference is due to distinct sample sizes (190 respiratory specimens in the previous report and ten in the present study) biasing the results, and/or to different underlying concomitant pulmonary diseases, in both Cuban and Spanish colonized populations. The low frequencies of SOD110C and SOD215T alleles are usually associated with more virulent PcP episodes.

The relationships between the most frequent P. jirovecii putative MLG detected in the different pools studied pointed out that clusters are mainly due to the geographic differences but also dependent on clinical characteristics of the populations studied (Fig. 2). MLG from pools corresponding to colonized patients from Portugal and Spain were grouped in the same cluster (cluster A). Also, the MLG from the pool from Cuba (infants infected with P. jirovecii) was included in this cluster but with a higher phylogenetic distance. The majority of the MLG detected in the Portuguese pools were clustered, corresponding to PcP cases (cluster B). Pneumocystis jirovecii from the USA and Mozambique were related in the same cluster because of the DHPS mutations present in both MLG (cluster C). However, within this cluster the MLG from Mozambique (infants infected with P. jirovecii) was considerably distant from the four major MLG detected in the pool from the USA (severe AIDS-related PcP cases).

The observation of different dihydropteroate synthase (DHPS) alleles in and between P. jirovecii populations carries epidemiological implications related to transmission patterns, sulfa drug exposure and geographical distribution of specific genotypes [14,15,17]. In general overview, patients living in the USA and Mozambique presented higher rates of DHPS mutants, which is clearly depicted in Fig. 1. This fact suggests that the circulation of P. jirovecii haplotypes may be potentially related to trimethoprim-sulfamethoxazole resistance in those geographical regions. The high frequencies of DHFR312T and SOD110C/SOD215T detected in patients with AIDS-related PcP from Portugal and the USA endorse the importance of these genetic variants of P. jirovecii, which were already associated with the virulence or severity of the disease (Table 3) [18,21]. Considering the present data, P. jirovecii with the haplotype DHFR312T/ SOD110C/SOD215T is likely to be associated with more severe AIDS-related PcP cases and high P. jirovecii burdens as demonstrated in the pools from patients with severe PcP from Portugal (PT2) and the USA.

Conclusion

Geographic location and clinical parameters of the groups of patients studied as well as the time-period in which the samples are obtained, were confirmed as determinant epidemiological factors of P. jirovecii infection. The present study demonstrates that the multifactorial approach to PcP studies is a powerful high-throughput methodology for large-scale screening of P. jirovecii SNP and epidemiological relevance. These results convey a more complete picture of the worldwide distribution of P. jirovecii haplotypes and assess their epidemiological impact in different geographic populations of patients.

Transparency Declaration

The present study was supported by Fundação para a Ciência e a Tecnologia (FCT) project PTDC/SAU-MIC/116716/2010 and Associação para a Investigação e Desenvolvimento da Faculdade de Medicina de Lisboa. Dr Huang was supported by National Institutes of Health (NIH) grants HL087713 and HL090335. This work was developed in the framework of the ‘Red Iberoamericana sobre Pneumocystosis’ (21/2T0450) of The Ibero-American Programme for Science, Technology and Development (CYTED).

References


© 2016 European Society of Clinical Microbiology and Infectious Diseases. Published by Elsevier Ltd. All rights reserved. CMI, 22, 566.e9–566.e19