Restriction Fragment Length Polymorphism Typing Demonstrates Substantial Diversity among *Pneumocystis jirovecii* Isolates

Chiara Ripamonti,1 Abigail Orenstein,a,b Geetha Kutty,1 Laurence Huang,1 Regina Schuhegger,4 Andreas Sing,4 Giovanna Fantoni,1 Chiara Atzori,1 Carol Vinton,1 Charles Huber,1 Patricia S. Conville,4 and Joseph A. Kovacs1

1Critical Care Medicine Department and 2Department of Laboratory Medicine, National Institutes of Health Clinical Center, National Institutes of Health, Bethesda, Maryland; 3Department of Medicine, San Francisco General Hospital, University of California San Francisco, San Francisco; 4Bavarian Health and Food Safety Authority (Bavarian LGL), Oberschleissheim, Germany; and 5Department of Infectious Diseases, Luigi Sacco Hospital, Milan, Italy

Better understanding of the epidemiology and transmission patterns of human *Pneumocystis* should lead to improved strategies for preventing *Pneumocystis* pneumonia (PCP). We have developed a typing method for *Pneumocystis jirovecii* that is based on restriction fragment length polymorphism (RFLP) analysis after polymerase chain reaction amplification of an ∼1300 base-pair region of the *msg* gene family, which comprises an estimated 50–100 genes/genome. The RFLP pattern was reproducible in samples containing >1000 *msg* copies/reaction and was stable over time, based on analysis of serial samples from the same patient. In our initial analysis of 48 samples, we found that samples obtained from different individuals showed distinct banding patterns; only samples obtained from the same patient showed an identical RFLP pattern. Despite this substantial diversity, samples tended to cluster on the basis of country of origin. In an evaluation of samples obtained from an outbreak of PCP in kidney transplant recipients in Germany, RFLP analysis demonstrated identical patterns in samples that were from 12 patients previously linked to this outbreak, as well as from 2 additional patients. Our results highlight the presence of a remarkable diversity in human *Pneumocystis* strains. RFLP may be very useful for studying clusters of PCP in immunosuppressed patients, to determine whether there is a common source of infection.

*Pneumocystis jirovecii* is an opportunistic fungus that causes pneumonia in immunocompromised patients, especially those with HIV infection, in whom it remains a major cause of morbidity and mortality [1–3]. Understanding the epidemiology of human *Pneumocystis* infection can be important in minimizing the clinical impact of this pathogen.

Recently, a number of molecular biologic methods have been developed to study the epidemiology, transmission patterns, and potential emergence of antimicrobial resistance of this organism [4–10]. At least 14 unique genetic loci have been evaluated using a variety of typing methods, including DNA sequence, single strand conformation polymorphism, and restriction fragment length polymorphism (RFLP). In general, genes that evolve very slowly, such as the mitochondrial small subunit ribosomal RNA (rRNA) and 5.8S rRNA genes, have been used in evolutionary studies or to examine genetic diversity. Genes that show greater diversity, such as the internal transcribed spacer region of the nuclear rRNA genes, have been used to evaluate case clustering and to look for evidence of direct person-to-person transmission. Genes that are targets of
therapeutic drugs, such as the dihydropteroate synthetase and dihydrofolate reductase genes, have been used to study the development of drug resistance.

The major surface glycoprotein is a Pneumocystis surface protein encoded by a multicopy msg gene family with an estimated 50–100 copies of related but unique msg genes per genome; only a single msg variant, which is present at a unique (single copy) expression site, the upstream conserved sequence, is expressed in a given organism [11–17]. We previously showed that a region in the intron of the upstream conserved sequence of P. jirovecii contains variable numbers of 10 base-pair (bp) tandem repeats that can be easily quantitated and used in epidemiologic studies [18]. More recently, we have demonstrated by RFLP analysis that the repertoire of msg genes (50–100 genes per genome) shows substantial variation among P. jirovecii isolates but not among Pneumocystis carinii (infecting rats) or Pneumocystis murina (infecting mice) isolates [19]. This suggests that RFLP analysis of msg genes could be used as an epidemiologic tool to investigate human Pneumocystis infection. The current study was undertaken to determine whether msg repertoire variability could be used as a typing method to distinguish among different human isolates of Pneumocystis.

MATERIALS AND METHODS

Patient specimens. A total of 48 samples with isolates of P. jirovecii, which included autopsy lung, bronchoalveolar lavage (BAL), sputum, and oral wash samples, were obtained from 40 patients from San Francisco, California (n = 8); other parts of the United States (n = 8); the Netherlands (n = 3) [18]; and Milan (n = 21) [20]. An additional 26 amplifiable samples were obtained from 22 patients from Germany [21]. Guidelines of the US Department of Health and Human Services were followed in the conduct of these studies.

DNA amplification of msg variable region. Genomic DNA was extracted using the QIAMP DNA mini Kit (Qiagen), according to the manufacturer’s instructions. Nested (initial studies) or semi-nested polymerase chain reaction (PCR) was used to amplify ∼1.3 kb of the msg variable region with use of the following primers: for the first round, GK 126, 5′-GTGTCGGCGG-GCGGT-3′ (corresponding to 2138–2151 bp of P. jirovecii msg upstream conserved sequence; GB#AF033209; for later studies) and GK 452, 5′-ACGATTGTATGGTCTGGT-3′ (corresponding to 2875–2893 bp of P. jirovecii msg HuMSG14). PCR was performed using HotStart Taq (Qiagen). The conditions were 15 min at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, and 4 min (for the first round) or 2 min (for the second round) at 72°C, with a final extension of 10 min at 72°C. msg Copy number was quantitated by a previously described real-time quantitative PCR assay [22].

Restriction enzyme treatment. PCR products were purified using QuickStep 2 PCR Purification Kit (Edge BioSystems), according to the manufacturer’s instructions. The purified products were digested with DraI, HindIII, or XbaI restriction enzymes for 5–6 h at 37°C. The digested products were analyzed on a 1.2% TBE agarose gel and were visualized by SYBR green staining (Molecular Probes). After transfer to a Nytran membrane (Whatman), the blot was hybridized with a digoxigenin-labeled DNA probe (PCR DIG Probe Synthesis Kit; Roche). The PCR probe (spanning ∼1.3 Kb) was an equal mixture of 4 products obtained by PCR amplification (as described above) of lung samples from 4 P. jirovecii–infected individuals. The hybridization signal was detected by chemiluminescence with use of alkaline phosphatase–conjugated anti-digoxigenin antibody and CDPStar (Roche). The results were recorded with a Luminescent Imager (Kodak Image Station 440CF; PerkinElmer).

Analysis of gels and blots. The gels and blots were analyzed using BioNumerics software (version 4.01; Applied Maths). The pattern of banding among different gels and/or blots was normalized using internal standards that were included in each run: Lambda/HindIII molecular weight markers for gels and a clinical sample (sample number 385) for blots. Molecular weights were assigned to the bands of the standards, and sample bands were identified manually. The Dice coefficient was used to analyze the similarity of the patterns of bands with a position tolerance of 1.9% [23]. The unweighted pair group method with average linkages was used by the BioNumerics software for cluster analysis. DNA samples with banding patterns with 100% similarity (Dice coefficient, 1) were considered to be identical.

RESULTS

RFLP assay development and reproducibility. We previously demonstrated substantial variation in RFLP patterns among a small number of P. jirovecii isolates by using analysis of the entire msg sequence (∼3200 nucleotides) from autopsy lung samples [19]. Because of difficulties in amplifying the entire sequence when using samples with lower organism loads (eg, induced sputum), we developed a semi-nested PCR that amplified an ∼1300 nucleotide region of msg with use of primers from conserved regions of msg (based on alignment of available P. jirovecii msg clones).

In preliminary studies, RFLP analysis of this shorter msg fragment also showed substantial variation among isolates. To investigate the reproducibility of this assay, we ran replicate PCRs using Pneumocystis DNA extracted from lung tissue and oral wash samples. The PCR products were digested with HindIII or DraI restriction enzymes and were analyzed by
agarose gel electrophoresis or Southern blotting. We found that the pattern was reproducible with use of lung tissue samples but not consistently with use of oral wash samples (data not shown), suggesting that the reproducibility of the RFLP pattern is dependent on organism load. By serial dilution of a single sample, we found that this assay lost reproducibility at <1000 msg copies per PCR (as determined by quantitative PCR [22]; data not shown). On the basis of these data, only samples with >1000 msg copies per PCR were considered to be reliably reproducible, and samples with <1000 msg copies were used with caution.

**RFLP pattern stability over time.** To investigate the stability of the RFLP pattern in samples obtained from 1 individual over time, we examined paired samples collected over varying periods. Samples collected at close time points (eg, <3 months apart) likely represent the same episode of PCP and likely would exhibit the greatest stability. We analyzed 16 samples from 8 individuals (10 BAL samples, 2 sputum samples, and 4 oral wash samples). The collection time between samples ranged from 1 day to 111 days. All BAL sample pairs (from patients 1 and 2) had an identical RFLP pattern (Figure 1), and the pair of sputum samples (from patient 3) had an identical RFLP pattern (Figure 1); however, there were differences in banding intensity among the paired isolates, which may represent variation in the proportions of cointecting P. jirovecii types. The oral wash sample pairs (from patients 4 and 5) (Figure 1) had highly similar patterns; however, for patient 4, the patterns were not identical, possibly because of low msg copy numbers. These data demonstrate that the RFLP pattern is stable over at least a period of days to weeks and that recurrent episodes of PCP (in patient 2) (Figure 1) can result from relapse rather than from reinfection with a new strain.

**RFLP analysis of multiple isolates.** We then performed RFLP studies to compare a larger number (n = 48) of Pneumocystis isolates that we had collected over time. Because RFLP analysis with Dra1 alone appeared sufficient to distinguish among isolates, these studies used only Dra1 in the initial analysis. Because samples needed to be run on different gels, a known sample (sample number 385) that would hybridize to the probe during Southern blotting, as well as commercial molecular weight standards, were included in each gel to allow comparison among different runs.

On the basis of visual examination, gels and blots showed an identical RFLP banding pattern only in samples collected from the same individual and a distinct RFLP pattern in samples obtained from different individuals. To allow comparison of samples run on different gels, we used BioNumerics software, with standardization using molecular weight markers (for gels) and sample 385 (for blots). Figure 2 shows a dendrogram created by BioNumerics software from different gels, after normalization using the internal standards. For occasional samples, there were differences in the precise clustering between gels and blots, which may be related to differences following hybridization with the probe specific for P. jirovecii (data not shown).

The dendrogram reveals that, in general, only samples obtained from the same patient showed 100% similarity. However, this analysis identified 4 Italian samples that formed 2 pairs, with each pair showing 100% similarity (samples 2999 and 3960 and samples 2428 and 7780). All the Italian samples were collected in the same city (Milan) during 1994–1999; however, because of unlinking, we cannot further analyze the patient data to determine whether these samples were collected from the same patient or at close time points. To explore this further, we digested 1 pair of samples (2999 and 3960) with another.

![Figure 1. Stability of the restriction fragment length polymorphism (RFLP) banding pattern in samples obtained over time. Representative gels (left) and corresponding blots (right) show the RFLP pattern of 2 samples (A and B) collected from 5 individuals at different times. Samples from patients 1 and 2 were bronchoalveolar lavage (BAL) samples that were obtained 15 and 111 days apart, respectively; samples from patient 3 were sputum samples that were obtained 10 days apart; and samples from patients 4 and 5 were oral wash samples that were both obtained 1 day apart. The samples from patient 2 were obtained during 2 separate episodes of Pneumocystis pneumonia. Although the samples from patients 4 and 5 had msg copy numbers below the cutoff of reproducibility of our assay (patient 4 B, 260 msg copies/reaction; patient 5 A and B, both 900 msg copies/reaction), the RFLP pattern was identical (patient 5) or highly similar (patient 4). All samples were digested with Dra1.](https://academic.oup.com/jid/article-abstract/200/10/1616/881526/ptl)
Figure 2. Similarity analysis of restriction fragment length polymorphism (RFLP) patterns (dendrogram derived by BioNumerics software from RFLP analysis of 37 samples following agarose gel electrophoresis). Eleven samples with <1000 msg copies per assay were excluded from this analysis. The percent similarity scale is shown above the dendrogram and is indicated by the numbers at the individual nodes. The samples grouped with the same color are from the same country: orange, Italy; green, the Netherlands; blue, the United States. Bronchoalveolar lavage samples that were collected from the same patient are indicated by a red box around the number. The green box indicates a sputum pair collected from the same patient. Five major clusters that were identified by the BioNumerics software are indicated by blue bars on the right and are numbered 1 to 5.

The Pneumocystis isolates that we analyzed were collected in the United States, Italy, and the Netherlands. Of interest, the samples collected in the same country clustered more closely to each other than to samples from other countries (Figure 2). In particular, the samples from Italy and the United States, which accounted for the majority of isolates, were not randomly intermingled.

Application of RFLP to investigate an outbreak of PCP.

Because of the substantial variability among Pneumocystis isolates from different patients and the stability of RFLP patterns...
within individuals over time, RFLP analysis appeared to provide a method for easily demonstrating whether isolates from a potential outbreak of PCP were identical. A recent study of an outbreak of PCP among renal transplant recipients in Germany provided molecular evidence, primarily by single nucleotide polymorphism analysis, that all patients were infected with a single *Pneumocystis* strain [21]. *Pneumocystis* DNA from 22 German patients was provided to us for RFLP analysis: 13 isolates were from patients linked to the outbreak, and 9 were unrelated to the outbreak. In 12 isolates previously identified as a single strain of *Pneumocystis* by single nucleotide polymorphism analysis, RFLP analysis showed an identical banding pattern (Figure 3A), although patient 8 (<1000 msg copies/assay) had an additional band not seen in the other 11 isolates. RFLP analysis with a second enzyme, Xba1, confirmed that the banding pattern was identical in all 12 isolates (data not shown). One additional isolate from the outbreak had a different banding pattern, but the sample contained <200 msg copies/assay. Of the 9 isolates from patients who were not previously linked to the outbreak, 6 showed a unique pattern, 2 showed a pattern identical to the outbreak pattern (Figure 3B), and 1 was similar but had <1000 msg copies/assay; RFLP analysis with Xba1 again confirmed these results (data not shown). Clinical history obtained after these results confirmed that all 3 of the latter patients were renal transplant recipients who had been seen at the same clinic as the other outbreak patients and who underwent bronchoscopy during the period of the outbreak. When the samples from Germany were included in the dendrogram, the non-outbreak-related samples did not cluster with the outbreak strain but tended to cluster with samples from Italy (data not shown).

**DISCUSSION**

We developed a reproducible and easy-to-perform method to type human *Pneumocystis* strains with use of RFLP analysis. By using this method, we demonstrated a remarkable diversity among human *Pneumocystis* strains: no 2 isolates from different patients showed an identical RFLP pattern, other than those from a cohort of German patients previously linked to a nosocomial outbreak of PCP. In contrast, samples from the same patient (that were obtained within 111 days of each other) showed an identical pattern. Thus, this method may be very useful for studying transmission patterns and potential outbreaks of PCP among immunosuppressed patients. Moreover, our data strongly support the previously published conclusions that the renal transplant recipients from Germany were infected with the same *Pneumocystis* strain, and we identified at least 2 additional renal transplant recipients who were likely to have been part of the same outbreak.

A strength of RFLP analysis is that, rather than examining a single or very limited number of nucleotide polymorphisms, as is the case with many available typing methods, it interrogates the entire msg repertoire of the *Pneumocystis* genome, which is estimated to include 50–100 genes, with ~40% of each msg of ~3200 nucleotides being evaluated in the RFLP analysis. Because there are multiple msg copies per genome and there is a high level of sequence conservation in short stretches across msg genes, it is likely that recombination in *Pneumocystis* can lead to rearrangements and establishment of unique msg repertoires, as we and others have previously shown [19]. The conservation in RFLP pattern among isolates from the same

![Figure 3](https://example.com/figure3.png)
patient, as well as the conserved pattern among a cohort of patients linked epidemiologically, suggest that recombination does not occur within a period of days to weeks. Previously, we revealed that the RFLP pattern in mouse and rat \textit{Pneumocystis} isolates obtained from inbred animals at 2 locations over a period of years were identical or highly similar. If we assume that human \textit{Pneumocystis} strains are biologically similar, it appears likely that repertoire evolution is not rapid and that the observed diversity is related to recombination that has evolved over many, perhaps thousands or millions of years.

One potential disadvantage of RFLP analysis is that it is unable to distinguish the contributions of individual isolates to the banding pattern in patients infected with >1 isolate. If only one of multiple strains is transmitted to a new host, however, the RFLP bands should be easily distinguishable as a subset of those in the first host. In addition, the reproducibility of RFLP analysis was lost in samples with <1000 \textit{msg} copies. This likely represents a sampling bias in a specimen with a low organism load that resulted in uneven distribution of \textit{msg} variants in different aliquots.

On the basis of the dendrogram analysis, isolates from samples obtained from patients from the same geographic area at approximately the same time did not show 100% similarity (other than 1 pair of Italian samples and those from the renal transplant recipients), suggesting that interhuman transmission among these patients did not occur to any significant extent. These data are in agreement with the results of a previous study [24] in which the authors showed, using PCR–single strand conformation polymorphism typing, that transmission of \textit{Pneumocystis} from patients with active PCP to susceptible persons is rare. However, outbreaks with the same strain can occur, as shown among the renal transplant recipients. It is noteworthy that samples collected in the same country clustered more closely with each other than with samples from other countries. This clustering may represent local strain variation. It is intriguing to speculate that host immune pressures at the population level (eg, human leukocyte antigen mediated) are driving the diversity of the \textit{msg} repertoire, as has been reported for HIV [25].

In summary, this study revealed a broad diversity in \textit{Pneumocystis} strains, provided a method for rapidly typing strains, and provided confirmatory evidence that an outbreak of PCP was caused by a single strain of \textit{Pneumocystis}. Larger studies using this approach are needed to better define the epidemiology of PCP and to determine whether any predominant strains, as defined by RFLP analysis, can be identified.

Acknowledgments

We thank Dr. Pieter J. A. Beckers for providing samples of \textit{Pneumocystis jirovecii} and allowing us to include them in this study.

References


