Genetic Variation in Pneumocystis carinii Isolates from Different Geographic Regions: Implications for Transmission


*Centers for Disease Control and Prevention, Atlanta, Georgia, USA; †University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; ‡University of California, San Francisco General Hospital, San Francisco, California, USA; §Emory University School of Medicine, Division of Infectious Diseases, Atlanta, Georgia, USA; ¶Veterans' Affairs Medical Center, Atlanta, Georgia, USA; #University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

To study transmission patterns of Pneumocystis carinii pneumonia (PCP) in persons with AIDS, we evaluated P. carinii isolates from patients in five U.S. cities for variation at two independent genetic loci, the mitochondrial large subunit rRNA and dihydropteroate synthase. Fourteen unique multilocus genotypes were observed in 191 isolates that were examined at both loci. Mixed infections, accounting for 17.8% of cases, were associated with primary PCP. Genotype frequency distribution patterns varied by patients' place of diagnosis but not by place of birth. Genetic variation at the two loci suggests three probable characteristics of transmission: that most cases of PCP do not result from infections acquired early in life, that infections are actively acquired from a relatively common source (humans or the environment), and that humans, while not necessarily involved in direct infection of other humans, are nevertheless important in the transmission cycle of P. carinii f. sp. hominis.

Pneumocystis carinii pneumonia (PCP) in humans, caused by the opportunistic fungal pathogen Pneumocystis carinii f. sp. hominis, is a frequent cause of illness in persons with AIDS. In recent years, the number of new cases of PCP has greatly diminished (1), largely as a result of highly active antiretroviral therapy and effective anti-Pneumocystis chemophrophylaxis, primarily with trimethoprim-sulfamethoxazole (TMP-SMZ). However, PCP is still the most common life-threatening AIDS indicator condition in the United States in patients whose CD4+ cell count has declined to <200 cells/µL for the first time (2).

The number of persons at risk, as measured by the number of HIV-infected patients with CD4+ cell counts <200 cells/µL, was recently estimated at around 182,000 (3).

Despite the frequency of PCP, tremendous gaps exist in understanding the basic biology and epidemiology of the causal agent. These gaps, which have been reviewed extensively (4-9), include lack of knowledge of population structure of P. carinii strains circulating among patients or in potential environmental reservoirs. The implications are important for understanding patterns of transmission and developing methods of intervention.

In the last 5 years, a substantial number of genes and gene fragments have been identified for potential use in analysis and characterization...
of \( P. \) carinii strains. These loci include the mitochondrial large subunit ribosomal RNA gene (\textit{mtlsurRNA}) (10-12) and the internal transcribed spacers of the nuclear ribosomal RNA array (12-14). The dihydropteroate synthase (DHPS) gene locus, which encodes a target for the anti-Pneumocystis drugs TMP-SMZ and dapsone, has also been cloned and sequenced (15). Substantial variation at this locus (16,17) suggests that the widespread use of antimicrobial chemoprophylaxis may be exerting selective pressure on \( P. \) carinii strains circulating in humans. Its polymorphism makes the locus not only potentially useful as a marker for changes in antimicrobial susceptibility levels, but also valuable for strain characterization and typing.

We examined polymorphism at two genetic loci of \( P. \) carinii isolates from persons with AIDS diagnosed in five U.S. cities. One locus, \textit{mtlsurRNA}, is involved in basic metabolic functions, and the other, DHPS, is the target of sulfone and sulfonamide antimicrobial drugs. We examined the population structure of \( P. \) carinii strains for information that would elucidate their patterns of transmission.

**Materials and Methods**

**Patient Samples**

Specimens used in the study were obtained from March 1995 to June 1998 during routine diagnostic procedures for HIV-infected patients hospitalized with PCP in Atlanta, Cincinnati, Los Angeles, San Francisco, and Seattle. A portion of the diagnostic specimen, either induced sputum or bronchoalveolar lavage, was preserved directly with an equal volume of absolute ethanol and stored at 4°C for DNA extraction.

**DNA Purification**

Specimens were divided into approximately 1-mL aliquots and centrifuged at 14,000 x g for 5 to 7 minutes. The resulting cell pellet was resuspended in 1 mL of phosphate-buffered saline (0.01M, pH 7.2) containing 1 mM EDTA (PBS-EDTA), washed twice in PBS-EDTA, centrifuged, and stored at -80°C for later DNA extraction. DNA was prepared by a commercial purification procedure (Wizard Genomic DNA Purification Kit, Promega, Madison, WI) in accordance with the product recommendations for DNA purification from blood. Final pellets were resuspended in 50 µL of TE (10 mM Tris, 1 mM EDTA, pH 7.2).

**Polymerase Chain Reaction (PCR) and DNA Sequencing**

PCR amplification was performed at two independent genetic loci. A 360-bp fragment was amplified from the \textit{mtlsurRNA} locus by using the published primers PAZ102E (5' - GAT GGC TGT TTC CAA GCC CA - 3') and PAZ102H (5' - GTG TAC GTT GCA AAG TAC TC - 3') (10). PCR conditions included a 94°C hot start for 5 minutes; followed by 35 cycles of a program consisting of 92°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds; followed by a termination step at 72°C for 5 minutes. The DHPS locus was amplified by a modification of a nested PCR procedure (16,17). In the first round of this PCR, the primers DHPS F1 (5' - CCT GGT ATT AAA CCA GTT ATT AAA CCA GTT TTG CC - 3') (S.R. Meshnick, pers. comm.) and DHPS B45 (5' - CAA TTT AAT AAA TTT CTT TCC AAA TAG CAT C - 3') (16) were used. In the second round, the primers DHPS A\textsubscript{HU}M (5' - GCG CCT ACA CAT ATT ATG GCC ATT TTA AAT C - 3') and DHPS BN (5' - GGA ACT TTC AAC TTG GCA ACC AC - 3') (16) were used. The conditions for the first round were 94°C for 5 minutes, followed by 35 cycles of 92°C for 30 seconds, 52°C for 30 seconds, 72°C for 60 seconds, and a termination step at 72°C for 5 minutes. The conditions for the second round of PCR were 94°C for 5 minutes, then 35 cycles of 92°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds, followed by a termination step at 72°C for 5 minutes.

For analysis of the PCR-generated fragments, 10 µL of each 50-µL PCR amplification product was examined by horizontal gel electrophoresis on 1% agarose gels. The remaining 40 µL of successfully amplified reactions was purified by a commercial purification procedure (Wizard PCR Purification Kit, Promega, Madison, WI) and suspended in 50 µL of TE for DNA sequencing; 5 µL to 10 µL of each purified product was sequenced directly by using dye terminator chemistry (ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, PE Applied Biosystems, Foster City, CA) according to the manufacturer’s protocol, with the DNA oligonucleotide primers used for PCR amplification. The PCR fragments were sequenced on an ABI 377 automated DNA sequencer (PE Applied Biosystems, Foster City, CA) according to manufacturer’s recommendations. The sequenced DNA fragments were analyzed by using Sequence Navigator v.1.0.1.
Researcch

Vol. 6, No. 3, May–June 2000

Emerging Infectious Diseases

Table 1. Genotype analysis of Pneumocystis carinii isolates at the mtlsurRNA locus

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Nucleotide position/identity</th>
<th>Frequency n=324 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85/C; 248/C</td>
<td>123 (38.0)</td>
</tr>
<tr>
<td>2</td>
<td>85/A; 248/C</td>
<td>119 (36.7)</td>
</tr>
<tr>
<td>3</td>
<td>85/T; 248/C</td>
<td>30 (9.3)</td>
</tr>
<tr>
<td>4</td>
<td>85/C; 248/T</td>
<td>19 (5.9)</td>
</tr>
<tr>
<td>mixed</td>
<td></td>
<td>33 (10.2)</td>
</tr>
</tbody>
</table>

Table 2. Genotype analysis of Pneumocystis carinii isolates at the DHPS locus

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Nucleotide (amino acid) position/identity</th>
<th>Frequency n=220 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>165 (55)/A (Thr)</td>
<td>68 (30.9)</td>
</tr>
<tr>
<td>2</td>
<td>171 (57)/C (Pro)</td>
<td>13 (5.9)</td>
</tr>
<tr>
<td>3</td>
<td>165 (55)/A (Thr)</td>
<td>5 (2.3)</td>
</tr>
<tr>
<td>4</td>
<td>171 (57)/T (Ser)</td>
<td>109 (49.5)</td>
</tr>
<tr>
<td>mixed</td>
<td></td>
<td>25 (11.4)</td>
</tr>
</tbody>
</table>

Statistical Analysis

Statistical analyses were performed with SAS software version 6.12. Logistic regression was used to examine the association between place of diagnosis and place of birth and genotype for mtlsurRNA and DHPS gene loci.

Results

Amplification with Specific Primers

PCR amplification with the selected primer sets gave consistent results in most P. carinii-positive samples. The mtlsurRNA primers amplified a 360-bp fragment in 223 samples from Atlanta, Los Angeles, San Francisco, and Seattle. The DHPS primer sets amplified a 300-bp fragment in 220 of these samples. Both genetic loci were successfully amplified for 191 samples. In the Cincinnati dataset, 101 samples were amplified for the mtlsurRNA locus. The DHPS site was not examined for this group of samples. All the amplified fragments were sequenced, aligned, and examined for genetic polymorphism.

General Observations on Genotype Frequency

Four unique genotypes were observed for each of the two genetic loci examined. At the mtlsurRNA locus, all the genotypes were distinguished on the basis of polymorphism at nucleotide positions 85 and 248. Genotypes 1 (85:C; 248:C) and 2 (85:A; 248:C) were the most common (Table 1), occurring at similar frequencies and together accounting for 74.7% of the 324 samples analyzed. Genotypes 3 (85:T; 248:C) and 4 (85:C; 248:T) accounted for 9.3% and 5.9%, respectively. At the DHPS locus, genotypes 1 and 4 accounted for 80.4% of all samples (Table 2). Genotypes 2 and 3 were both relatively uncommon, seen in 5.9% and 2.3% of the samples analyzed, respectively. All the mutations seen at this locus were nonsynonymous changes resulting in amino acid substitutions; no other polymorphism was observed. Genotype 1 was the designation used to refer to the sequence defined by a threonine at position 55 and a proline at position 57. Genotype 2 referred to an alanine at 55 and a proline at 57. Genotype 3 referred to a threonine at position 55 and a serine at 57, and genotype 4 to an alanine at position 55 and a serine at position 57.

When the results at both genetic loci were combined, 14 unique multilocus genotypes of 16 possible combinations were observed in 191 samples for which both genes could be amplified and sequenced. The four most common multilocus genotypes accounted for 61.7% of all genotypes. The most common multilocus genotypes consisted of combinations of the most common genotypes at each individual locus. No genetic linkage of specific genotypes from the two loci was observed.

Coinfection with multiple P. carinii strains could be detected in 33 (10.2%) of 324 samples typed at the mtlsurRNA locus and 25 (11.4%) of 220 typed at the DHPS locus. When the genotypes were considered together, 34 (17.8%) of 191 samples represented coinfections with multiple genotypes. When these samples were analyzed according to patient history, 21 (17.6%) of 119 samples from patients with no history of previous PCP were coinfected with multiple strains, by typing at the mtlsurRNA locus. In contrast, none of the samples from 41 patients with PCP history had multiple genotypes (p = 0.002, Fisher’s exact test). A similar trend was observed at the DHPS locus but was not statistically significant.
Genotype Frequency Distribution Patterns by Geographic Location

Genotype frequencies in PCP isolates differed by city (Figures 1, 2). At the mtl-surRNA locus (Figure 1), the genotype distribution was significantly different (chi-square test; \( p = 0.001 \)), with the ratio of genotype 1 to genotype 2 ranging from 0.7 in San Francisco to 1.8 in Cincinnati. Analysis of the mtl-surRNA genotype frequencies at each city showed that genotype distributions also differed significantly or borderline significantly, with \( p \) values from 0.001 to 0.08, with the exception of Los Angeles. Comparisons involving Los Angeles were nonsignificant, perhaps as a result of small sample size (\( n = 15 \)).

We used logistic regression to investigate whether the geographic distribution was associated more with place of birth or with place of diagnosis. We grouped places of birth and diagnosis into either “east” or “west,” using the Mississippi River as the dividing line. Data from Cincinnati were not included, since place-of-birth information was not available there. None of the four mtl-surRNA genotypes was significantly associated with place of birth when the data were adjusted for place of diagnosis. However, genotypes 2 and 4 were significantly associated with place of diagnosis, when the data were adjusted for place of birth (\( p = 0.002, p = 0.05 \), respectively); the association of genotype 1 with place of diagnosis was borderline significant (\( p = 0.08 \)).

The overall genotype distribution at the DHPS gene locus (Figure 2) also differed significantly (Fisher’s exact test; \( p = 0.049 \)) for the various cities. The ratios of genotype 4 to genotype 1 ranged from 0.97:1 in Atlanta to 3.13:1 in San Francisco. No isolates from Cincinnati were analyzed at the DHPS locus.

Logistic regression analysis of the DHPS genotypes by place of diagnosis and place of birth showed results similar to those for the mtl-surRNA genotypes. Genotypes 1 and 4 were associated with place of diagnosis (\( p = 0.001, p = 0.002 \), respectively) when data were adjusted for place of birth. In contrast, place of birth was not associated with any DHPS genotype when data were adjusted for place of diagnosis.

Discussion

P. carinii, once thought to be a protozoon but now regarded as an ascomycetelike fungus (4,19,20), has been associated with human disease since the 1940s. Despite intense efforts to understand this important disease agent, lack of a suitable means of propagation has complicated study of basic biology and epidemiology. Similarly, latency and reactivation (versus recent acquisition) have not been resolved. PCR and other molecular methods have improved understanding of genetic variability and host specificity (11,12,14,20,21), as well as the cause of recurrent infections in persons with AIDS (8,22-24).

Both an indirect (environmental) source and a direct (person-to-person) source have been proposed as modes of transmission of P. carinii in humans. Three primary observations reported in this study address infection sources and transmission patterns: geographic variation in
Observations suggest that similar to that among adults with AIDS. These prevalence of PCP among these infants was very PCP episodes at 3 to 6 months of age; the perinatally acquired HIV infection (25) who had observations of primary PCP in infants with subsequent infections do not necessarily represent relapses. The second line of study comprises patients (22,23), which suggests that infections in adults are acquired later than the first few months or years of life and that any latency has natural limits. Two independent lines of study offer a context for these observations and suggest that PCP is an actively acquired infection. The first line of study comprises molecular genetic analyses of P. carinii strains from adult AIDS patients with recurrent PCP. These analyses have shown that different P. carinii genotypes are detected on subsequent PCP episodes in a substantial proportion of patients (22,23), which suggests that infections in adults are actively acquired and that subsequent infections do not necessarily represent relapses. The second line of study comprises observations of primary PCP in infants with perinatally acquired HIV infection (25) who had PCP episodes at 3 to 6 months of age; the prevalence of PCP among these infants was very similar to that among adults with AIDS. These observations suggest that P. carinii is common in the environment, consistent with the suggestion that the organism is easily acquired.

While the association with geographic distribution is consistent with recent acquisition of clinical infections, the actual source of the infection is not known. The rat P. carinii model suggests direct (animal-to-animal) transmission (5,26,27); however, if person-to-person transmission does occur, it has not been shown to be of epidemiologic significance (28,29). PCR amplification of P. carinii DNA from spore traps from various sites supports the possibility of an environmental source (30-33), but no specific plant, animal, or soil source has been identified. Identification of a specific environmental reservoir for P. carinii f. sp. hominis, if one exists, could elucidate disease transmission and improve prevention efforts.

The rate of coinfection with multiple P. carinii strains by multilocus typing was 17.8% (34 of 191). Of the 191 patients for whom both genes were typed, 20 (10.5%) were coinfected at the mtsurRNA locus and 19 (9.9%) at the DHPS locus. Five patients tested positive for multiple strains concurrently at both loci. While the sensitivity to detect coinfections was similar for both genes, concordance between the two genes was very low in detecting coinfections with multiple strains in any particular isolate—only four genotypes can be detected at each individual locus. When the loci are considered together, however, the sensitivity is greatly increased, because the number of possible genotypes increases from 4 to 16, 14 of which were observed in this study.

The 17.8% coinfection rate is within the 10% to 30% range reported by other investigators using DNA sequencing-based approaches (14,22,34). Although coinfection rates as high as 69% have been reported with single-strand conformation polymorphism analysis (35), these differences can probably be resolved by two considerations. The first is the sensitivity of the selected locus, which is a function of the genetic variability or evolutionary rate of the locus (i.e., the more variable the locus, the greater its sensitivity to detect a different genotype). In this study, mtsurRNA and DHPS displayed a high degree of genetic conservation, with only four alleles detected at each locus, all of which have been observed (11,16,17). Because these are relatively slowly evolving genes, the sensitivity to detect coinfections with multiple strains is expected to be less than in a faster evolving locus. The second consideration is PCP prophylaxis. Correlation between primary PCP and coinfection with multiple strains suggests that patients may be exposed to multiple P. carinii strains over extended periods and harbor short-lived, latent infections. Consequently, patients who have not been treated are more likely to have been exposed (and to harbor) multiple P. carinii strains. On the other hand, patients who have been treated for PCP at least once and are taking secondary prophylaxis are less likely to become reinfected. This hypothesis is consistent with indications that P. carinii is ubiquitous in the environment.
Research

and exposure in humans is commonplace (22,23). A short latency period has been suggested (36) and is consistent both with confections’ correlating with primary PCP and allelic frequency distributions’ correlating with the place of diagnosis but not the place of birth.

The question of latency and reactivation versus recent acquisition is of great importance as it relates to prevention. If the preponderance of infections in humans results from activated latent infections, chemoprophylaxis is the only method of preventing disease. If, however, infections are actively acquired, identifying specific sources of infection would lead to other methods of prevention and reduce dependence on antimicrobial agents.

The most important clinical implication of the polymorphism observed specifically at the DHPS locus relates to the emergence of possible antimicrobial resistance; however, the observation of different DHPS alleles in P. carinii populations also has implications for transmission routes and patterns and the possibility of person-to-person transmission.

In human P. carinii isolates, four distinct genotypes have been reported at the DHPS locus (16,17). All four of these DNA base changes result in amino acid substitutions at an important structural position, in an otherwise very highly conserved gene (17,37). Genotype 1 in our study corresponds to the “wild-type” genotype, as defined by the only allele observed in P. carinii from any other mammalian species (16,17). Genotype 2 is a point mutation that results in a threonine-to-alanine substitution at amino acid position 55. Genotype 3 is also a point mutation, resulting in a proline-to-serine substitution at amino acid position 57. Genotype 4 is a double mutant, which has alanine at position 55 and serine at position 57. This particular mutation has rarely been associated with failure of both TMP-SMZ treatment (38) and prophylaxis (17).

The presence of polymorphism at two different positions, both associated with amino acid substitutions, at a gene locus that otherwise lacks variability, is itself indicative of pronounced selective pressure. Such a degree of selection is not itself an indication of resistance, but certainly suggests that resistance may be emerging. Mutations in this same region of the DHPS molecule have been correlated with specific antimicrobial resistance to sulfa drugs in a number of other microorganisms, including Plasmodium falciparum (39), Streptococcus pneumoniae (40), Streptococcus pyogenes (41), Escherichia coli (42), and Neisseria meningitidis (43).

The two single mutations are uncommon by themselves (5.9% for genotype 2 and 2.3% for genotype 1), yet 50% of all isolates have the double mutant, which suggests that, alone, neither mutation is highly selected, but together they pose a very strong selective advantage. If humans are dead-end hosts for P. carinii, how can such a pronounced degree of polymorphism be explained? Person-to-person transmission may be essential to allow genetic selection to occur, resulting in this polymorphism.

Geographic variation in allelic frequency detected at the mtlsurRNA locus is not unexpected, even though other studies with smaller sample sizes failed to observe differences (11). Mitochondrial DNA sequence data have been highly useful in detecting intraspecific differences between populations of diverse organisms (44,45). Perhaps more unexpected, however, was the detection of geographic variation at the DHPS locus. Unlike mtlsurRNA, because it encodes a gene product that is the target of the primary anti-P. carinii drug TMP-SMZ, the DHPS locus is assumed to be subject to intense selection pressure. Consequently, this selection might be expected to override any potential variation from geographic separation or geographic patterns to reflect TMP-SMZ exposure patterns. The relative frequency of the double mutant genotype was much higher in samples from the West Coast, particularly San Francisco, than in samples from Atlanta (Figure 2). The reason for this difference is not obvious because in preliminary multivariate analysis conducted with samples from Atlanta and San Francisco, when the data are controlled for sulfa exposure, place of diagnosis is still the most significant factor influencing the frequency of the double mutant genotype (data not shown). Explanations for this observation are being evaluated.

An overall frequency of approximately 50% for the DHPS double mutant genotype is somewhat higher than that reported in other recent studies (46,47). The reason for this observation is also unclear. One possible explanation is that the samples were collected more recently (March 1995 to June 1998) and therefore reflect a sulfa-induced increase in
frequency of the double mutant. Most patients were from San Francisco and Atlanta, the two populations that showed the greatest difference in genotype frequency at the DHPS locus. Sample collection in Atlanta began in 1995, but in San Francisco the earliest samples were from 1997. When we stratified the patients in Atlanta by date, 1995-96 versus 1998-99, we saw no significant difference in the frequency of mutant genotypes. Similarly, when we stratified data for San Francisco by date, 1997 versus 1998-99, we saw no differences. Consequently the frequency of the mutation did not change during this study period, nor does the variation observed according to geography appear to be confounded by the dates when specimens were collected.

Conclusions
The pattern of allelic variation differed at each of the cities where samples were obtained, and this variation correlated with the place of diagnosis but not with the place of birth. Coinfection with multiple *P. carinii* genotypes was associated with primary rather than secondary PCP. The position 55/57 double mutant accounted for 50% of all genotypes of the DHPS locus examined. These observations suggest the following possibilities, which contradict much of the current opinion on the epidemiology and transmission of PCP: Most cases of PCP are not a result of infections acquired very early in life; infections are actively acquired from a relatively common source (humans or the environment); and humans, while not necessarily involved in direct infection of other humans, are nevertheless important in the transmission cycle of *P. carinii* f. sp. hominis.

Acknowledgments
The authors thank Lori Ballard and Joan Turner for technical assistance; Dan Coley, Sue Binder, and Mary Bartlett for critically reading the manuscript; Steven Meshnick for designing the specific DNA oligonucleotide primers used for PCR amplification; and Brian Holloway and the staff at the NCID Biotechnology Core Facility for preparing the primers.

Informed consent was obtained from all patients. Human experimentation guidelines of the U.S. Department of Health and Human Services and those of participating institutions were followed in the conduct of this research.

References
43. Fermer C, Kristiansen BE, Skold O, Swedberg G. Sulfonamide resistance in Nesseria meningitidis as defined by site-directed mutagenesis could have its origin in other species. J Bacteriol 1995;177:4669-75.