Phylogeny of *Rickettsia rickettsii* Genotypes Associated with Fatal Rocky Mountain Spotted Fever


Infectious Diseases Pathology Branch and Rickettsial Zoonoses Branch, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; Núcleo de Anatomia Patológica, Centro de Patologia do Instituto Adolfo Lutz, São Paulo, Brazil; Departamento de Vigilância em Saúde, Secretaria Municipal de Saúde de Campinas, São Paulo, Brazil

Abstract. Rocky Mountain spotted fever (RMSF), a tick-borne zoonosis caused by *Rickettsia rickettsii*, is among the deadliest of all infectious diseases. To identify the distribution of various genotypes of *R. rickettsii* associated with fatal RMSF, we applied molecular typing methods to samples of DNA extracted from formalin-fixed, paraffin-embedded tissue specimens obtained at autopsy from 103 case-patients from seven countries who died of RMSF. Complete sequences of one or more intergenic regions were amplified from tissues of 30 (29%) case-patients and revealed a distribution of genotypes consisting of four distinct clades, including the Hlp clade, regarded previously as a non-pathogenic strain of *R. rickettsii*. Distinct phylogenetic patterns were identified when composite case-patient and reference strain data were mapped to the state and country of origin. The phylogeny of *R. rickettsii* is likely determined by ecological and environmental factors that exist independently of the distribution of a particular tick vector.

INTRODUCTION

Rocky Mountain spotted fever (RMSF), a tick-borne, bacterial zoonosis caused by infection with *Rickettsia rickettsii*, is among the deadliest of all known infectious diseases. Historical mortality estimates from the sentinel epidemic of RMSF in the Bitterroot Valley of western Montana are staggering: 88 (75%) of 118 well-documented cases of RMSF that occurred here during the late 19th and early 20th centuries ended in death. During 1900–1909, 92 RMSF deaths were recorded in Ravalli County alone, representing an annual mortality rate of nearly 1/1,000 individuals. Despite the recognized severity of RMSF in Montana, the earliest reports of this disease also documented far milder forms in other western states, suggesting to clinicians and investigators that an extraordinarily virulent strain of *R. rickettsii* existed in the Bitterroot Valley.

During the next 50 years, investigators obtained multiple isolates of *R. rickettsii* from humans and from various species of wildlife and ticks that exhibited a wide spectrum of virulence in experimentally infected animals. Rickettsiologists subsequently evaluated many of these isolates by using immunoblot and immunofluorescence techniques to determine if observations from animal experimentation might correlate to qualitative differences in one or more proteins among recognized strains. These methods provided relatively limited resolution among strains and it was not until the beginning of the 21st century when whole-genome sequencing provided investigators an effective tool to dissect the genetic heterogeneity of *R. rickettsii*. At least three molecular typing strategies have been developed that provide strain separation using nucleotide polymorphisms. One of these systems, based primarily on availability and on the relative completeness of national surveillance for fatal RMSF. Fisher’s exact two-tailed test was used for comparisons of gender, race, and age between the case-patients in this series and the same series were matched by gender, age, state, and year of death to information from case report forms (CRFs) to assess completeness of national surveillance for fatal RMSF. Fisher’s exact two-tailed test was used for comparisons of gender, race, and age between the case-patients in this series and the same categorical data summarized by CRFs during 1983–2007.

The Wilcoxon rank-sum test was used to compare survival times of case-patients in this series with survival times presented for 84 RMSF decedents from the Bitterroot Valley during 1885–1903.

DNA extraction, amplification, and sequencing. When available, a minimum of two blocks of different FFPE tissues
from each case-patient were processed. A 16-μm section was cut from each tissue block, deparaffinized in xylene, washed twice in absolute ethanol, and air-dried. Deparaffinized tissues were digested at 56°C in Proteinase K and digestion buffer (Qiagen, Valencia, CA) for a minimum of 12 h. The DNA was extracted using the QIAamp DNA Mini Kit (Qiagen) and extracted using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) for a minimum of 12 h. The DNA was digested at 56°C, twice in absolute ethanol, and air-dried. Deparaffinized tissues were eluted in a final volume of 100 μL. The integrity of each extract was evaluated by a multiplex polymerase chain reaction (PCR) assay targeting a 198-base-pair (bp) segment of the human glyceraldehyde-3-phosphate dehydrogenase gene and a 514-bp segment of the human β-globin gene. Samples were tested for segments of the Rickettsia genome by using multiple PCR assays (Table 1), including a broad-range, genus-specific real-time assay for the citrate synthase (csp) gene, a real-time assay targeting a Rickettsia-specific hypothetical protein, and nested PCR assays targeting the RR0155-rpmB, cspA-ksgA, and RR1240-tlc5 intergenic regions of R. rickettsii. The locus spo0J-abcT1 was additionally assessed for a case-patient infected with an Hlp genotype by using a nested PCR assay. Gel-purified PCR products were sequenced using a PRISM BigDye Terminator Cycle 3.1 sequencing kit (Applied Biosystems, Foster City, CA) and an ABI 3130xl genetic analyzer (Applied Biosystems), or a Dye Terminator Cycle Sequencing Quick Start Kit (Beckman Coulter, Indianapolis, IN) and a GenomeLab GeXP Genetic Analysis System (Beckman). DNA sequence analyses. Sequences were assembled using Sequencher 5.1 (Gene Codes Corporation, Ann Arbor, MI) and aligned using ClustalW and MEGA5. MEGA5 was used to infer the evolutionary history using the maximum likelihood method based on the Tamura-Nei model. One thousand bootstrap replicates were used to estimate the likelihood of each tree.

## RESULTS

### Case-patient demographics

Tissue specimens were obtained from 103 case-patients with an IHC-established diagnosis of fatal RMSF during 1981–2013 (Table 2). Case-patients originated from seven countries; however, ~90% were from the United States. **Table 2** Demographic and clinical characteristics of 103 case-patients with fatal Rocky Mountain spotted fever from North, Central, and South America, and molecular features of infecting genotypes of *Rickettsia rickettsii*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>United States (N = 63)</th>
<th>Mexico (N = 10)</th>
<th>Panama (N = 2)</th>
<th>Costa Rica (N = 3)</th>
<th>Colombia (N = 3)</th>
<th>Brazil (N = 19)</th>
<th>Argentina (N = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number male (%)</td>
<td>34 (54)</td>
<td>4 (40)</td>
<td>0</td>
<td>2 (67)</td>
<td>2 (67)</td>
<td>16 (84)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Median age in years (range)</td>
<td>21 (0–73)</td>
<td>8 (0–50)</td>
<td>13 (4–22)</td>
<td>12 (8–65)</td>
<td>21 (20–65)</td>
<td>32 (6–54)</td>
<td>8 (3–11)</td>
</tr>
<tr>
<td>Median days from onset to death (range)</td>
<td>8 (5–12)</td>
<td>8 (4–18)</td>
<td>6 (4–9)</td>
<td>7 (6–8)</td>
<td>6 (6–7)</td>
<td>7 (4–12)</td>
<td>9 (8–14)</td>
</tr>
<tr>
<td>Number with data for genotypic analysis (%)</td>
<td>22 (35)</td>
<td>4 (40)</td>
<td>1 (50)</td>
<td>0</td>
<td>1 (33)</td>
<td>1 (5)</td>
<td>1 (33)</td>
</tr>
<tr>
<td><strong>Clade(s) identified</strong></td>
<td>I, II, IV</td>
<td>I</td>
<td>III</td>
<td>–</td>
<td>III</td>
<td>III</td>
<td>III</td>
</tr>
</tbody>
</table>

*Clade designations defined by previous analysis of historical reference strains."
the United States, Mexico, and Brazil. The majority of case-patients (61%) were from the United States, representing 30 states and the District of Columbia. Nineteen (18%) of the total case-patients originated from the states of São Paulo, Espírito Santo, Rio de Janeiro, and Minas Gerais in Brazil, and 10 (10%) from the states of Baja California and Coahuila in Mexico. Argentina, Costa Rica, and Columbia were each represented by 3 case-patients, and 2 case-patients originated from Panama.

Case-patients were represented by 42 females and 61 males who ranged in age from 2 months to 73 years. The median age was 20 years; 47 (46%) were < 18 years of age, and 25 (24%) were < 6 years of age, including three infants. A duration of illness was available for 92 (89%), and the median survival time was 8 days (range, 4–18 days); 67 (73%) died within 6–9 days of illness. The overall survival time for case-patients in this series was marginally shorter (P value = 0.02) when compared with the aggregate survival time reported for a large series of decedents from the Bitterroot Valley who died of RMSF more than 100 years ago (Figure 1).1

Case report forms were identified for 33 (52%) of the United States case-patients. These included 9 (64%) of the 14 cases that occurred during 1981–1991, 12 (46%) of 26 during 1992–2002, and 12 (52%) of 23 during 2003–2013. Among the case-patients from the United States, no significant differences were noted between the percentages of male (54%) and non-white (18%) individuals evaluated in our study and the corresponding percentages (59% and 18%, respectively) among RMSF deaths documented by CRF surveillance data during 1983–2007.22,23 However, our analysis included significantly more United States case-patients < 10 years of age (40% versus 15%, P value < 0.0001), and significantly fewer case-patients ≥ 40 years of age (29% versus 63%, P value < 0.0001) when compared with the ages of United States RMSF decedents reported to CDC during the same approximate period.

**PCR amplification of nucleic acids.** Samples of DNA were extracted from a total of 221 FFPE tissue blocks. The most commonly represented tissues, in order of frequency, were lung, spleen, liver, kidney, central nervous system, adrenal gland, skin, heart, pancreas, thyroid gland, testis, thymus gland, bone marrow, lymph node, and skeletal muscle. Segments of the glyceraldehyde-3-phosphate dehydrogenase and β-globin genes were amplified successfully from tissues of 80 (78%) and 29 (28%) case-patients, respectively. One or more tissue extracts were positive by the broad-range, Rickettsia-specific and R. rickettsii-specific real-time assays for 75 (73%) and 55 (53%) case-patients, respectively. Complete sequences of one or more intergenic regions were obtained from tissues of 30 (29%) case-patients, comprising RR0155-rpmB for 30 (29%), cspa-ksgA for 19 (18%), and RR1240-tlc5 for 20 (19%). All 3 loci were sequenced from the tissues of 17 (16%) case-patients (Table 2). In general, success at amplifying the targeted sequences was related inversely to the age of the tissue block. No genotyping data were obtained from any case-patients who died before 1991, and complete data from 1 or more intergenic regions were obtained from only 3 (6%) of 46 who died before 2003. Regardless of the age of the sample, fewer sequences were amplified from FFPE tissues of patients from countries other than the United States. Accordingly, genotyping data were available from 19 (83%) of 23 United States case-patients who died during 2003–2013, but from only 8 (24%) of 34 non-United States case-patients from this same period.

Sequence analyses of the RR0155-rpmB and RR1240-tlc5 loci revealed an Hlp genotype of *R. rickettsii* from the tissue of one United States case-patient. Although no sequence was obtained for the cspa-ksgA locus, an additional 148-bp product was amplified from the spotI-abcTI locus, which contained the 7-bp insertion (GTATAAA) unique to the Hlp genotype of *R. rickettsii*.19 The case-patient was a 48-year-old Caucasian man with alcohol dependence who succumbed to an illness of at least 7 days duration, characterized by fever to 40.5°C, headache, malaise, and delirium. Laboratory assessment revealed severe thrombocytopenia, leukocytosis, and markedly elevated levels of creatinine phosphokinase. Histological evaluation of tissues obtained at autopsy revealed lymphohistiocytic perivascular and interstitial infiltrates in lungs, kidneys, heart and central nervous system. An immunohistochemical staining method demonstrated SFG rickettsiae in endothelial cells of small and medium sized vessels in each of these tissues (Figure 2). No specific history of tick bite was documented; however, the case-patient reported multiple exposures to rodents, rabbits, and deer.

**Phylogenetic analyses.** Phylogenetic analysis of RR0155-rpmB and RR1240-tlc5 sequences, amplified from one or more tissues of 30 and 20 case-patients, respectively, revealed a distribution of genotypes comprising four distinct clades (Figure 3), as defined by a previous analysis of reference strains of *R. rickettsii*.19 An analysis of the complete and concatenated typing loci from the tissues of 17 case-patients corroborated the distribution of these clades (Figure 4). Distinct phylogeographic patterns were identified when the composite case-patient data were mapped to the country and state of origin and these patterns were consistent with the geographical distributions of archival strains isolated from ticks, animals, and humans (Figure 5).19 Clade I was restricted to the eastern, central, and southwestern United States; clade II, to the Pacific Northwest and West Coast regions of North America; and clade III, to Central and South America.
A single infection with an Hlp genotype, comprising clade IV, was identified in a case-patient from the northern Great Plains of the United States.

**DISCUSSION**

By analyzing molecular characteristics of *R. rickettsii* DNA obtained from archival FFPE tissues of case-patients with fatal RMSF, we identified natural infections in human hosts with genotypes from each of the four currently recognized clades of *R. rickettsii*. The distribution of these genotypes among multiple countries in the Western Hemisphere exhibited phylogeographical associations that corroborated a recent genetic analysis of reference isolates acquired from various sources during several decades.\(^\text{19}\) Collectively, these data identify two predominant clades in the United States and northern Mexico, and one in Central and South America. Previous genotypic analyses of *R. rickettsii* have included many strains cultivated from ticks and wildlife, and many that were serially passaged in guinea pigs, yolk sacs, or cultured cells, including some with > 20 known passages.\(^\text{17,19,21}\) Our study was unique in that it characterized genotypes of *R. rickettsii* obtained exclusively from patients with fatal disease, obviating genetic or phenotypic changes, including virulence, that may result from repeated *in vitro* cultivation.\(^\text{13,32}\)

To our knowledge, this is the first description of RMSF caused by an Hlp genotype of *R. rickettsii*. The identification of disease in a human caused by this genotype follows recent discoveries of other rickettsioses in the Americas caused by *Rickettsia* species considered non-pathogenic for many decades.\(^\text{33}\) For more than 60 years rickettsiologists characterized the Hlp genotype as a relatively non-virulent variant of *R. rickettsii* and generally irrelevant in discussions of RMSF in humans.\(^\text{15,16,34}\) The first characterized isolates of the Hlp genotype were obtained from *Haemaphysalis leporispalustris* ticks removed from cottontail rabbits collected in the Bitterroot Valley during 1948. Guinea pigs infected with Hlp strains exhibited milder disease than those infected with a classically virulent strain of *R. rickettsii*, including longer incubation periods, shorter febrile intervals, diminished scrotal pathology, and most notably, the absence of fatal outcomes in the Hlp-infected animals.\(^\text{10}\) Subsequent reports largely validated these observations with additional isolates obtained from *H. leporispalustris* ticks collected in Pennsylvania and Virginia during 1940; nonetheless, a few deaths were noted among some guinea pigs infected with these Hlp isolates.\(^\text{11}\) and later investigations in the eastern United States and Costa Rica identified Hlp isolates that produced high fevers, scrotal necrosis and substantial mortality in guinea pigs.\(^\text{35,36}\) Limited genetic analyses of a Costa Rica Hlp isolate reveals complete identity with several
United States isolates of *R. rickettsii*, including the Hlp #2 and Sheila Smith strains. Collectively, these data reinforce the observation that the severity of disease caused by a particular strain of *R. rickettsii* in a non-human host cannot be extrapolated routinely to the severity of disease in a human host, and that additional molecular analyses are needed to better elucidate genetic determinants of virulence in *R. rickettsii*.

At least seven different species of hard ticks are considered potential vectors of *R. rickettsii* and the geographical distributions of these vectors are well characterized at a continental scale. Using data from previous genotyping analyses, investigators have suggested that the distribution of a particular *R. rickettsii* genotype is likely determined by the geographical distribution of a corresponding tick vector. In this scheme,
clade I strains from the eastern United States associate with *Dermacentor variabilis* ticks, whereas clade II strains from the western states associate with *Dermacentor andersoni* ticks. Nonetheless, data from this study and others suggest that the phylogeography of *Rickettsia rickettsii* is more likely characterized by ecological and environmental factors that exist independently of the distribution of a specific tick vector. From our evaluation, we identified a clade II genotype in a California case-patient who was infected several hundred miles outside of the geographical range of *D. andersoni*. Another exception is exemplified by the involvement of *Rhipicephalus sanguineus* ticks in separate outbreaks of RMSF caused by clade I strains of *R. rickettsii* in Arizona and in Coahuila, Mexico, and data herein, despite natural infections with *R. rickettsii* in at least four potential vector species across this region.

The results of this evaluation indicate that fatal RMSF may be caused by any of the four recognized clades of *R. rickettsii*. It is axiomatic that any strain capable of causing death in a human host is a virulent strain; however, the data from this study are insufficient to assign relative degrees of virulence among the individual clades. Historically, isolates of *R. rickettsii* obtained from the Bitterroot Valley were considered particularly lethal and it is possible that an uniquely virulent genotype of *R. rickettsii* existed in western Montana during the early 1900s; nonetheless, the aggregate survival time of case-patients in our series, representing >50 separate locations across the continental United States and six countries of Latin America, was in fact marginally shorter than that described previously for patients in the Bitterroot Valley more than a century ago. Comparisons of strain virulence must also be considered in the context of various recognized, and possibly other, as-yet-unidentified host factors that

**Figure 4.** Dendrogram comprising concatenated sequences of RR0155-rpmB, cspA-ksgA, and RR1240-tlc5 sequences from formalin-fixed and paraffin-embedded tissue sections of 17 patients with fatal Rocky Mountain spotted fever. Previously evaluated reference strains of *Rickettsia rickettsii*, isolated predominantly from patients with fatal RMSF, are marked by asterisks. Bootstrap values (1,000 replicates) are shown above the nodes. *Rickettsia conorii* represents the out-group. The scale bar corresponds to the number of steps.
affect the severity of RMSF. Patients who are ≥ 40 years of age and African–American males with glucose-6-phosphate dehydrogenase deficiency have consistently higher case-fatality rates, however, neither of these cohorts were represented disproportionately in our series. In fact, among United States case-patients, significantly more young children and fewer persons ≥ 40 were represented when compared with national surveillance statistics for fatal RMSF in the United States during 1983–2007. Chronic alcohol abuse is another risk factor that may increase the severity of RMSF, in a manner similar to its demonstrated association with cases of fatal Mediterranean spotted fever, caused by the closely related bacterium, *Rickettsia conorii*. Although information regarding alcohol use was not available for most case-patients in our series, it is important to note that the single Hfp-associated death occurred in a patient with alcohol dependency.

This investigation was limited by its analysis of specimens obtained exclusively from RMSF decedents for whom an autopsy was performed, and for whom these specimens were subsequently submitted to CDC or Institute Adolfo Lutz for diagnostic evaluation. Perhaps most importantly is that our analysis was limited by the restrictions inherent to any molecular evaluation of formalin-fixed tissue. Preservation of tissues in formalin causes nucleic acids to crosslink and fragment, which characteristically limits PCR amplification to shorter segments of DNA from these specimens. This was reflected by our inability to amplify a 514-bp segment of the human β-globin gene from tissues of 72% of the case-patients. Multiple other factors, including the purity and pH of the formalin solution, the duration of fixation, and the age of the fixed tissue in the paraffin block can hamper or entirely inhibit amplification of a desired genomic target. These
restrictions prohibited genetic analyses of *R. rickettsii* from the tissues of 70% of the infected case-patients, and a complete set of three typing loci were available for only 16% of the decedents in this series; nonetheless, all data to date reveal considerable fidelity among the individual loci, i.e., the clade defined by a specific RR0155-rpmB genotype is associated invariably with the same *cspA-ksgA* and RR1240-tlc5 genotypes.\textsuperscript{18–20}

Only 27% of the total genotypic information was obtained from Central and South American countries. It is possible that subsequent evaluations of clinical isolates or well-preserved DNA samples will reveal greater genetic heterogeneity of *R. rickettsii* in these regions. These analyses could be particularly important in Brazil, where the national case-fatality rate of RMSF from 1989 to 2008 was ~25%,\textsuperscript{49} considerably higher than contemporary case-fatality estimates of 5–10% in the United States. Indeed, some investigators have suggested that a highly virulent strain of *R. rickettsii* predominates in Brazilian cases of RMSF.\textsuperscript{50}

During the early 20th Century, physicians in western Montana were powerless to prevent the unrelenting and frequently terminal course of RMSF. In 1902, a clinician from the Bitterroot Valley opined:

> The name spotted fever I consider a good one, for this reason: to the laity it bodes grave responsibility. We well know that certain diseases afflicting a patient to a profound degree are inconsistent with life. Many of these cases terminating fatally with the best skill and untiring energy an intelligent physician can give them, we then feel the burden is somewhat lightened, and the responsibility of life is shared by family and friends if they are prepared for the inevitable.\textsuperscript{51}

During the 1930s until the mid-1940s, ~75 to 140 RMSF deaths were recorded annually in the United States.\textsuperscript{42,52} The United States case-fatality rates for RMSF diminished considerably during the second half of the 20th century as physicians discovered effective antibiotic therapies and supportive medical treatment of this disease; since 1995, fewer than 10 fatal cases of RMSF have been reported annually in the United States.\textsuperscript{24} Nonetheless, the actual burden of fatal RMSF remains difficult to assess, and contemporary estimates of the case-fatality rate are confounded by underreporting of laboratory-confirmed deaths.\textsuperscript{22} In this series, approximately half of the laboratory-confirmed United States case-patients were not matched to a CRF, and were therefore absent from national surveillance estimates of fatal RMSF. These data corroborate previous capture-recapture analyses of national mortality statistics for fatal RMSF, which estimated that ~65% of the RMSF deaths that occurred during 1983–2007 were not reported for national surveillance.\textsuperscript{23,24}

More than a century following its discovery, RMSF continues to kill otherwise healthy persons throughout the Western Hemisphere every year. With the recognition that infection with any of the known clades of *R. rickettsii* may result in a fatal outcome, additional investigations are needed to better characterize specific host factors that place infected patients at greatest risk for the most severe forms of this disease.

Acknowledgments: We thank Theodore Tzianabos (CDC), for producing the antiserum used in the immunohistochemical assay, and for providing access to several of the archival cases evaluated in this study: Onyekachuwu Nwanwko, for assistance with cataloguing the archival paraffin blocks; Marcus Nashelsky (University of Iowa Carver College of Medicine, Iowa City, Iowa), Kimberly Goble (Spearfish Regional Hospital, Spearfish, South Dakota), Sonia Montiel (San Diego County Health Department, San Diego, California), and Joanna Regan, Steven Waterman, Wun-Ju Shieh, Clifton Drew, Jana Ritter, Atis Muhenlebachs, and Sherif Zaki, (CDC) for their assistance in the clinical characterization and pathologic assessment of many of the case-patients evaluated in this series. This work was funded by the U.S. Department of Health and Human Services. The findings and conclusions are those of the authors and do not necessarily represent the official position of the U.S. Department of Health and Human Services.

Authors’ addresses: Christopher D. Paddock, F. Scott Dahlgren, and Sandra E. Karpathy, Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, Atlanta, GA, E-mails: cd9@cdc.gov, iot0@cdc.gov, and evu2@cdc.gov. Amy M. Denison, Lindy Liu, and Brigid C. Bollweg, Infectious Diseases Pathology Branch, Centers for Disease Control and Prevention, Atlanta, GA, E-mails: crk6@cdc.gov, fuz3@cdc.gov, and ftv6@cdc.gov. R. Ryan Lash, Traveler’s Health Branch, Centers for Disease Control and Prevention, Atlanta, GA, E-mail: gnk6@cdc.gov. Cristina T. Kanamura, Fabiana C. Pereira dos Santos, and Roosevelt Brasil Martins, Instituto Adolfo Lutz, Núcleo de Apoio à Pós Graduação e Pesquisa, São Paulo, São Paulo, Brazil, E-mails: criskanamura@yahoo.com.br, fabianapc@yahoo.com, and roosebrasili@hotmail.com. Rodrigo N. Angerami, Secretaria Municipal de Saúde de Campinas, Departamento de Vigilância em Saúde, Campinas, São Paulo, Brazil, E-mail: rodrigo.angerami@gmail.com.

REFERENCES


