CONCISE COMMUNICATION

Correlation between Respiratory Syncytial Virus Genotype and Severity of Illness

Richard A. Martinello,[1,2] Martin D. Chen,[3,a] Carla Weibel,[1] and Jeffrey S. Kahn[1]

Departments of [1]Pediatrics and [2]Internal Medicine, Division of Infectious Diseases, and [3]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut

Respiratory syncytial virus (RSV) causes seasonal outbreaks of respiratory tract infections, but the viral factors associated with virulence remain unknown. To determine whether RSV genotype correlated with severity of illness, isolates were characterized by phylogenetic analysis of the RSV G gene, and a composite score was used to quantify severity of illness. During the 1998–1999 and 1999–2000 winter seasons, 137 subgroup A and 84 subgroup B isolates were identified. The severity of illness caused by subgroup A isolates did not differ from that caused by subgroup B isolates ($P = .086$). However, the GA3 clade was associated with significantly greater severity of illness, compared with clades GA2 ($P = .004$) and GA4 ($P = .016$). In a subpopulation of patients ⩽24 months old who had no known risk factors for severe RSV disease, clade GA3 was again associated with greater severity of illness, compared with clade GA2 ($P = .018$). Severity of RSV infection is associated with RSV genotype.

Respiratory syncytial virus (RSV) is a major pathogen causing seasonal outbreaks of upper and lower respiratory tract infections in children and adults [1, 2]. By 2 years of age, >90% of children have serologic evidence of RSV infection [3]. In the United States, >100,000 children are hospitalized annually as a result of RSV infection, and these rates have been increasing since the early 1980s [1]. Comorbid conditions, such as congenital heart disease, chronic lung disease, and prematurity, have been associated with greater severity of RSV disease [4], but the viral factors associated with virulence and severity of disease are still incompletely understood.

RSV isolates are separated into subgroups A and B by antigenic and genetic characteristics [5]. This dimorphism is primarily due to variation within the G glycoprotein, a surface-expressed glycoprotein putatively associated with attachment of the virus. Each season, many distinct strains of both RSV subgroups circulate, and the predominant endemic strains vary from season to season [5, 6]. Previous studies that compared severity of illness of patients infected with subgroup A and B either have revealed no significant clinical differences or have concluded that infection with subgroup A isolates is associated with more-severe illness [7].

Whether specific strains within each subgroup are associated with the severity of illness remains unclear. Hall et al. [8] observed an association between RSV strains (grouped by immunological methods) and intensive care admissions. Fletcher et al. [9], using restriction fragment–length polymorphism (RFLP) analysis of viral cDNA to define RSV strains, described an association between a virus "group" and severity of disease. Isolate groupings derived via immunological methods or RFLP analysis may not be sensitive enough to distinguish individual strains. Phylogenetic analysis of the G gene has been used to define RSV genotype [6] and may possess greater sensitivity than other methods. The G gene is ideal for this analysis, because highly variable regions are present within the gene. We sought to determine whether RSV genotype is associated with disease severity.

Methods

Collection of clinical isolates and viral culture. All clinical specimens that tested positive for RSV antigen by direct immunofluorescence assay (DFA) were collected from the Yale–New Haven Hospital Clinical Virology Laboratory (New Haven, CT) during the 1998–1999 and 1999–2000 RSV seasons (November through April). These specimens were cultured for RSV on HEp-2 cells as described elsewhere [7].

RNA extraction, reverse-transcription (RT) polymerase chain reaction (PCR), and sequencing. Primers used for the amplification of the RSV G gene were as follows (primer G/C clamps are under-

Received 28 March 2002; revised 21 May 2002; electronically published 9 August 2002.

Presented in part: 39th annual meeting of the Infectious Diseases Society of America, San Francisco, 25–28 October 2001 (abstract 410).

Collection of clinical isolates and data was approved by the Yale University Human Investigation Committee.

Financial support: National Institutes of Health (grants HD-27757 to J.S.K. and AI-07517 to R.A.M.); American Lung Association and American Lung Association of Connecticut (grant RG-042-N to J.S.K.).

[a] Present affiliation: University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.

Reprints or correspondence: Dr. Jeffrey S. Kahn, Dept. of Pediatrics, Div. of Infectious Diseases, Yale University School of Medicine, PO Box 208064, New Haven, CT 06520-8064 (jeffrey.kahn@yale.edu).

The Journal of Infectious Diseases 2002; 186:839–42

0022-1899/2002/18606-0016$15.00

lined): forward primer 5′-CCGCGGGTTCTGGCAATGATAATC-TCAAC-3′, corresponding to the transmembrane domain of the G gene (genome location 4821–4842, subgroup A [reference strain RSV A Melbourne, Australia/2/1961; GenBank accession no. M74568], genome location 4823–4844, subgroup B [reference strain RSV B/ West Virginia/14617/1985; GenBank accession no. AF013254]); subgroup A reverse primer 5′-GCCGCGTGTATAATTCATAAACCT-TGGTAG-3′ (genome location 5991–6015); and subgroup B reverse primer, 5′-GGGGCCCCGCGGCCGCGCATTAATAGCAAGA-GTTAGGAAG-3′ (genome location 5698–5721). RNA was extracted, using the RNeasy Mini Kit (Qiagen) according to the manufacturer's recommendations, from the cell lysates of cultures that displayed cytopathic effects consistent with RSV. RT reactions were done with Moloney murine leukemia virus RT (New England BioLabs), and PCR was performed with HotStarTaq DNA polymerase (Qiagen) according to the manufacturer's specification. PCR amplification cycles were as follows: 95°C for 15 min followed by 35 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, and a final 10-min cycle at 72°C. Sequencing was done on Applied Biosystems 377 DNA automated sequencers at the W. M. Keck Biotechnology Resource Laboratory, Yale University School of Medicine (New Haven, CT).

Phylogenetic analysis. Sequence alignment was performed with the DNASTAR (DNASTAR) and PILEUP (Genetics Computer Group) software. Sequence comparisons were performed on a 243-nt segment of the G gene proximal hypervariable sequence. The PHYLIP software package (available at http://evolution.genetics.washington.edu/phylip.html) was used for phylogenetic analysis and clade assignment. Maximum likelihood phylogenetic trees were constructed by use of the PHYLIP program DNAML with a default transition-to-transversion ratio of 2.0. Five hundred bootstrap datasets were created using the PHYLIP program SEQBOOT. Clades with bootstrap values of <70% were considered insufficiently robust and were analyzed as part of their parent subset [10]. RSV G gene sequences available from GenBank were included in the phylogenetic analysis to compare the genetic characteristics of local isolates with geographically and temporally distinct isolates [6].

Clinical severity score (CSS). A CSS was adapted from the severity score described by Walsh et al. [7]. Two points were assigned if the patient required mechanically ventilatory support during the illness, and 1 point was assigned for each of the following: hospital admission, hospitalization for ⩾5 days, oxygen saturation ⩽87%, and use of supplemental oxygen. Therefore, the CSS ranged from 0 to 6.

Chart review. Medical records from patients who had positive results of DFA for RSV were abstracted and scored by a blinded reviewer. All subjects included in the study had positive results of DFA for RSV. Subjects who were coinfected with other respiratory viruses were excluded from the analysis. A subpopulation of otherwise healthy children ⩽24 months of age who lacked underlying comorbidities (i.e., gestation ⩽36 weeks, history of wheezing or reactive airway disorder, bronchopulmonary dysplasia, cystic fibrosis, cyanotic congenital heart disease, primary or secondary immunodeficiencies, and history of clinically significant gastroesophageal reflux/aspiration) potentially predisposing them to RSV illness of greater severity was identified [4, 11]. Children who had a history of a lower respiratory tract illness within the preceding 2 months, who had a history of a previous RSV infection, who had undergone surgery with general anesthesia in the preceding 2 weeks or had received intravenous immunoglobulin, RSV intravenous immunoglobulin, or palivizumab during the preceding 6 months were excluded from this subpopulation. Subjects >24 months old were not included in this subpopulation, because these patients were likely to have had previous exposure to RSV, which could have altered the severity of the current RSV illness [3].

Statistical analysis. The Wilcoxon rank sum test was used for analysis of continuous variables, and Fisher's exact test was used for analysis of categorical variables. All comparisons were 2-tailed. Parametric methods were not used, because CSS distributions were non-Gaussian and the sample sizes were unequal. The Bonferroni correction was used to adjust the P value threshold for statistical significance when multiple comparisons were made in which $\alpha = .05/k$, with k equal to the number of comparisons in which $P \leqslant .05$. All calculations were performed with SAS (version 6.12; SAS Institute) and Microsoft Excel Office 2000.

Results

During the 1998–1999 and 1999–2000 seasons, 355 patients with RSV infection were identified, and their clinical records were reviewed. The median age was 7 months; 78% were ⩽24 months old, and 5% were ⩾60 years. Of the 355 patients, 278 (78%) were hospitalized. RSV culture results were positive for 249 (70%) of 355 specimens. Subgroup assignments were made for 225 of the 249 culture-positive samples: 139 were subgroup A viruses, and 86 were subgroup B viruses. Four patients (2 with subgroup A and 2 with subgroup B) had concurrent infection with other respiratory viruses (2 with adenovirus and 2 with influenza viruses) and were excluded. Of the 137 remaining subgroup A isolates, 63 were segregated into the GA2 clade, 14 into the GA3 clade, and 53 into the GA4 clade [6]. No isolates from clade GA1 were identified during the study period. Clade assignments were not made for subgroup B isolates, because analysis revealed insufficiently robust bootstrap values of <70% [10]. Representative nucleotide sequences were submitted to GenBank (accession nos. AF510227–AF510315).

Comparison of subgroup A and B CSS distributions showed a trend toward an association between subgroup A and greater severity of illness ($P = .086$) (tables 1 and 2). Within the subgroup A isolates, clade GA3 was associated with significantly greater severity of illness, compared with clades GA2 ($P = .004$) and GA4 ($P = .016$). A trend was also seen toward an association between clade GA3 and greater severity of illness, compared to the B subgroup ($P = .079$). After adjustment for multiple comparisons, an association between clade GA3 and a significantly greater severity of illness, compared with clades GA2 and GA4, remained (table 2).

We sought to determine whether the association between clade and severity of illness was also present in a subpopulation of patients ⩽24 months old who lacked comorbidities predisposing them to severe RSV disease. Of 221 patients infected with a virus for which the subgroup had been defined, 107 were

Table 1. Distribution of CSSs, according to RSV subgroup/clade, among 221 patients infected with RSV and a subpopulation of 107 children ≤24 months old who lacked underlying comorbidities potentially predisposing to RSV illness of greater severity.

	All patients		Children without predisposing comorbidities	
Subgroup, clade	No. of patients	Mean CSS ± SD	No. of children	Mean CSS ± SD
A	137	1.2 ± 1.4	64	1.1 ± 1.5
GA2	63	1.0 ± 1.3	29	0.8 ± 1.1
GA3	14	2.2 ± 1.7	7	2.0 ± 1.4
GA4	53	1.1 ± 1.4	24	1.4 ± 1.8
B[a]	84	1.4 ± 1.4	43	0.9 ± 1.1

NOTE. CSSs, clinical severity scores; RSV, respiratory syncytial virus.
[a] Clade assignments were not made for subgroup B isolates, because analysis revealed insufficiently robust bootstrap values of <70%.

≤24 months old and did not meet risk-factor criteria. In the subpopulation analysis, clade GA3 was again associated with significantly greater severity of illness, compared with clade GA2 ($P = .018$) (tables 1 and 2), and clade GA3 was still associated with greater severity, compared with subgroup B ($P = .032$). No significant difference in severity of illness was noted between subgroups A and B ($P = .590$). After adjustment for multiple comparisons, the difference in severity between clades GA3 and GA2 remained statistically significant ($P = .018$) (table 2).

Our investigation of the association between genotype and severity of illness depended on the propagation of RSV in cell culture. To address the possibility that this introduced bias, we compared clinical data between RSV DFA–positive, culture-positive patients and RSV DFA–positive, culture-negative patients. No difference was noted in the mean CSS for the culture-positive group (1.3) and the culture-negative group (1.4) ($P = .42$).

Discussion

Previous studies have suggested that virus lineages within subgroup A are associated with greater illness severity. These studies distinguished virus lineages on the basis of immunological methods [8] or RFLP analysis [9]. Definition of virus group using immunological methods is limited, because several distinct virus strains may express the same antigenic domain. Likewise, "groups" of viruses sharing RFLPs may contain multiple strains with several distinct genetic polymorphisms. To better differentiate specific strains of RSV, we categorized isolates on the basis of phylogenetic analysis of genomic sequences. Our results suggest that a specific virus genotype within subgroup A is associated with greater severity of illness.

Because host factors play a major role in RSV pathogenesis, we also chose to analyze a subgroup of patients ≤24 months old who had no underlying risk factors, other than age, for severe RSV disease. Most children are infected with RSV by the age of 2 years, and RSV-positive children >2 years of age, therefore, likely would be experiencing a reinfection with RSV and might be protected against some strains through previous encounters with the virus. When the comparative analysis was performed using data from this subpopulation, the CSS differences between clades GA3 and GA2 remained statistically significant. Because this subpopulation is relatively homogenous and without known comorbidities, it provides a measure of severity of illness with fewer potential confounding factors.

In the respiratory virus seasons that we studied, clade GA3 was associated with greater virulence than the other clades identified. Our observations may be explained by shifting clade-specific immune status within the community. Because the predominant clades shift from year to year [6], severity of illness may be related to the prevalence of persons naive to specific clades within the population. Continued observation and analysis of additional seasons of RSV infection will be required to determine whether the association of greater severity of illness with clade GA3 infection persists or is solely transient.

Study of the pathogenesis of RSV infections is hampered by the lack of an adequate in vitro or animal model. Observations of genetic variants of RSV in cell culture and in animal studies do not necessarily coincide with those found in clinical isolates. For example, deletion of the short hydrophylic *(SH)* gene has no effect on viral growth in vitro and causes only slight attenuation in the animal model [12]. However, this gene is highly conserved in clinical isolates [13], which suggests that *SH* may

Table 2. Comparisons of the associations between mean CSS and RSV subgroup or clade among 221 patients with RSV infection and a subpopulation of 107 children ≤24 months old who lacked underlying comorbidities potentially predisposing to RSV illness of greater severity.

Cohort, subgroup/clade (mean CSS) vs. subgroup/clade (mean CSS)	P[a]
All patients	
A (1.2) vs. B (1.4)	.086
GA3 (2.2) vs. GA2 (1.0)	.004[b]
GA3 (2.2) vs. GA4 (1.1)	.016[b]
GA3 (2.2) vs. B (1.4)	.079
GA4 (1.1) vs. GA2 (1.0)	.724
B (1.4) vs. GA2 (1.0)	.041
B (1.4) vs. GA4 (1.1)	.153
Children without predisposing comorbidities	
A (1.1) vs. B (0.9)	.590
GA3 (2.0) vs. GA2 (0.8)	.018[c]
GA3 (2.0) vs. GA4 (1.4)	.241
GA3 (2.0) vs. B (0.9)	.032
GA4 (1.4) vs. GA2 (0.8)	.522
B (0.9) vs. GA2 (0.8)	.723
B (0.9) vs. GA4 (1.4)	.504

NOTE. CSS, clinical severity score; RSV, respiratory syncytial virus.
[a] Calculated with the 2-tailed Wilcoxon rank sum test.
[b] $P \leq .017$ was considered to be significant; the Bonferroni adjustment was used when k (the no. of comparisons for which $P \leq .05$) = 3.
[c] $P \leq .025$ was considered to be significant; the Bonferroni adjustment was used when k (the no. of comparisons for which $P \leq .05$) = 2.

play an essential role in the pathogenesis of RSV infection in humans.

The relationship between virus genotype and pathogenesis has been described for other viral diseases. The molecular basis for the highly pathogenic influenza virus strain H5N1 is a cleavable domain in the hemagglutinin gene and a point mutation in the *PB2* gene [14]. Among individuals infected with hepatitis C virus, response to therapy is dependent on virus genotype [15], and the cardiovirulence of coxsackievirus B3 has been localized to the 5′ untranslated region of the genome [16].

In conclusion, our study demonstrates that clade GA3 was associated with greater severity of illness, compared with other clades isolated during the 2 seasons we studied. RSV G gene sequences can be used as a genetic marker for virulence, because genetic recombination is not known to occur in paramyxoviruses. Description of the association between RSV genotype and severity of illness is a step toward the identification of virulence and/or genetic factors that play a role in pathogenesis.

Acknowledgments

We are indebted to Marie Landry, David Ferguson, and the staff of the Clinical Virology Laboratory, Yale–New Haven Hospital, for assistance in collection of clinical specimens; Ann Marie Quinn, for her expertise and assistance in the phylogenetic analysis; and Eugene D. Shapiro, for his thoughtful and critical review of the manuscript.

References

1. Shay DK, Holman RC, Newman RD, Liu LL, Stout JW, Anderson LJ. Bronchiolitis-associated hospitalizations among US children, 1980–1996. JAMA **1999**; 282:1440–6.
2. Zambon MC, Stockton JD, Clewley JP, Fleming DM. Contribution of influenza and respiratory syncytial virus to community cases of influenza-like illness: an observational study. Lancet **2001**; 358:1410–6.
3. Parrott RH, Kim HW, Brandt CD, Chanock RM. Respiratory syncytial virus in infants and children. Prev Med **1974**; 3:473–80.
4. Wang EE, Law BJ, Stephens D. Pediatric Investigators Collaborative Network on Infections in Canada (PICNIC) prospective study of risk factors and outcomes in patients hospitalized with respiratory syncytial viral lower respiratory tract infection. J Pediatr **1995**; 126:212–9.
5. Anderson LJ, Hierholzer JC, Tsou C, et al. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. J Infect Dis **1985**; 151:626–33.
6. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. J Gen Virol **1998**; 79:2221–9.
7. Walsh EE, McConnochie KM, Long CE, Hall CB. Severity of respiratory syncytial virus infection is related to virus strain. J Infect Dis **1997**; 175: 814–20.
8. Hall CB, Walsh EE, Schnabel KC, et al. Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children. J Infect Dis **1990**; 162:1283–90.
9. Fletcher JN, Smyth RL, Thomas HM, Ashby D, Hart CA. Respiratory syncytial virus genotypes and disease severity among children in hospital. Arch Dis Child **1997**; 77:508–11.
10. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol **1993**; 42:182–92.
11. Navas L, Wang E, de Carvalho V, Robinson J. Improved outcome of respiratory syncytial virus infection in a high-risk hospitalized population of Canadian children. Pediatric Investigators Collaborative Network on Infections in Canada. J Pediatr **1992**; 121:348–54.
12. Bukreyev A, Whitehead SS, Murphy BR, Collins PL. Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse. J Virol **1997**; 71:8973–82.
13. Chen MD, Vazquez M, Buonocore L, Kahn JS. Conservation of the respiratory syncytial virus *SH* gene. J Infect Dis **2000**; 182:1228–33.
14. Hatta M, Gao P, Halfmann P, Kawaoka Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science **2001**; 293:1840–2.
15. Poynard T, Marcellin P, Lee SS, et al. Randomised trial of interferon α2b plus ribavirin for 48 weeks or for 24 weeks versus interferon α2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. International Hepatitis Interventional Therapy Group. Lancet **1998**; 352: 1426–32.
16. Dunn JJ, Chapman NM, Tracy S, Romero JR. Genomic determinants of cardiovirulence in coxsackievirus B3 clinical isolates: localization to the 5′ nontranslated region. J Virol **2000**; 74:4787–94.