



DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU



HARVARD LIBRARY
Office for Scholarly Communication

Within-Host Whole-Genome Deep Sequencing and Diversity Analysis of Human Respiratory Syncytial Virus Infection Reveals Dynamics of Genomic Diversity in the Absence and Presence of Immune Pressure

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters

Citation	Grad, Y. H., R. Newman, M. Zody, X. Yang, R. Murphy, J. Qu, C. M. Malboeuf, J. Z. Levin, M. Lipsitch, and J. DeVincenzo. 2014. "Within-Host Whole-Genome Deep Sequencing and Diversity Analysis of Human Respiratory Syncytial Virus Infection Reveals Dynamics of Genomic Diversity in the Absence and Presence of Immune Pressure." <i>Journal of Virology</i> 88 (13) (April 16): 7286–7293. doi:10.1128/jvi.00038-14.
Published Version	doi:10.1128/JVI.00038-14
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:25123559
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://

Within-Host Whole-Genome Deep Sequencing and Diversity Analysis of Human Respiratory Syncytial Virus Infection Reveals Dynamics of Genomic Diversity in the Absence and Presence of Immune Pressure

Yonatan H. Grad,^{a,b} Ruchi Newman,^c Michael Zody,^c Xiao Yang,^c Ryan Murphy,^d James Qu,^c Christine M. Malboeuf,^c Joshua Z. Levin,^c Marc Lipsitch,^{a,e} John DeVincenzo^{d,f,g}

Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA^a; Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA^b; Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA^c; Department of Pediatrics, University of Tennessee School of Medicine, Memphis, Tennessee, USA^d; Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA^e; Department of Microbiology, Immunology, and Molecular Biology, University of Tennessee School of Medicine, Memphis, Tennessee, USA^f; Children's Foundation Research Institute, LeBonheur Children's Hospital, Memphis, Tennessee, USA^g

ABSTRACT

Human respiratory syncytial virus (RSV) is the leading cause of lower respiratory tract disease in infants and young children and an important respiratory pathogen in the elderly and immunocompromised. While population-wide molecular epidemiology studies have shown multiple cocirculating RSV genotypes and revealed antigenic and genetic change over successive seasons, little is known about the extent of viral diversity over the course of an individual infection, the origins of novel variants, or the effect of immune pressure on viral diversity and potential immune-escape mutations. To investigate viral population diversity in the presence and absence of selective immune pressures, we studied whole-genome deep sequencing of RSV in upper airway samples from an infant with severe combined immune deficiency syndrome and persistent RSV infection. The infection continued over several months before and after bone marrow transplant (BMT) from his RSV-immune father. RSV diversity was characterized in 26 samples obtained over 78 days. Diversity increased after engraftment, as defined by T-cell presence, and populations reflected variation mostly within the G protein, the major surface antigen. Minority populations with known palivizumab resistance mutations emerged after its administration. The viral population appeared to diversify in response to selective pressures, showing a statistically significant growth in diversity in the presence of pressure from immunity. Defining escape mutations and their dynamics will be useful in the design and application of novel therapeutics and vaccines. These data can contribute to future studies of the relationship between within-host and population-wide RSV phylogenetics.

IMPORTANCE

Human RSV is an important cause of respiratory disease in infants, the elderly, and the immunocompromised. RSV circulating in a community appears to change season by season, but the amount of diversity generated during an individual infection and the impact of immunity on this viral diversity has been unclear. To address this question, we described within-host RSV diversity by whole-genome deep sequencing in a unique clinical case of an RSV-infected infant with severe combined immunodeficiency and effectively no adaptive immunity who then gained adaptive immunity after undergoing bone marrow transplantation. We found that viral diversity increased in the presence of adaptive immunity and was primarily within the G protein, the major surface antigen. These data will be useful in designing RSV treatments and vaccines and to help understand the relationship between the dynamics of viral diversification within individual hosts and the viral populations circulating in a community.

Human respiratory syncytial virus (RSV) is the leading cause of severe viral respiratory disease in infants and young children (1) and is a significant cause of morbidity and mortality in the elderly (2) and immunocompromised (3, 4). Reinfection can take place throughout life, although in older children and healthy adults infection is usually associated with milder disease (5, 6). Population-wide molecular epidemiology studies have shown that there are multiple cocirculating RSV genotypes (where genotypes have been defined by the sequence of the variable region of the attachment [G] protein) that undergo a gradual replacement in dominant genotypes over successive seasons (7–9). Studies of RSV evolution in response to selective pressures have demonstrated escape mutations *in vitro* in the G protein (10, 11), and analyses of clinical collections of RSV sequences have found population-genetic evidence suggesting positive selection within the G protein (12–14). These studies support the hypothesis that the variation in circulating RSV genotypes reflects response to im-

une pressure. However, little is known about the extent of viral diversity over the course of an individual infection or the effect of immune status on viral diversity. Additionally, studies to date have focused almost exclusively on variations within small segments of the G protein, as these appear to be the most variable

Received 7 January 2014 Accepted 7 April 2014

Published ahead of print 16 April 2014

Editor: A. García-Sastre

Address correspondence to Yonatan H. Grad, ygrad@hsph.harvard.edu.

M.L. and J.D. contributed equally to this work.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JVI.00038-14>.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00038-14

regions of the genome. As the G protein is a surface protein primarily responsible for virus attachment to host cells and an immunomodulatory secreted form also exists (15), understanding the natural diversity of the G protein in the absence of specific immune pressure within a host is important. Likewise, understanding the diversity of the rest of the genome and the relationship of this variation to immune pressure may aid in development of therapeutics and guide vaccine targets and strategies.

Recent technological advances in deep sequencing have established a platform with which we are able to generate full-genome assemblies without the aid of a reference and to investigate the diversity of viral subpopulations over the course of individual infections. Deep sequencing studies of chronic RNA viral infections, such as HIV infections, have expanded our understanding of viral evolutionary dynamics in response to immunity and drug resistance (16, 17). Applying these same technologies to acute, self-limited infections with RNA viruses such as RSV can aid in understanding viral pathogenesis and the role of immune pressure in generation of viral sequence diversity. The assessment of the relationship between immune pressure and viral diversity and adaptation will allow for deeper understanding of the origins and dynamics of viral population diversity (18).

To study the within-host diversity of RSV and the role of immune pressure, we investigated the viral populations from an RSV-infected infant with severe combined immune deficiency (SCID) who subsequently underwent a bone marrow transplant (19). The infant developed progressive respiratory symptoms over the month prior to hospitalization. Subsequently, nasal aspirate samples were collected prospectively and viral loads quantified under a research protocol designed to study RSV pathogenesis. Treatment with 4 infusions of palivizumab had no clinical impact, and the nasal RSV load remained unchanged. He underwent bone marrow transplantation without prior myeloablative conditioning. Notably, the patient's father was the donor, and he had lived with the patient and had symptoms consistent with RSV infection 2 months prior to the transplant. At approximately 10 days posttransplant (termed D +10), viral load declined and became undetectable. However, the patient developed respiratory distress. Infusion of RSV immunoglobulin (Ig), palivizumab, and immunomodulatory therapy for possible graft-versus-host disease was associated with clinical improvement, as well as resurgence in nasal RSV load. The rise in absolute lymphocyte count and CD8 cell count started around day 12 posttransplant (D +12) and likely represented proliferation of donor-derived effector cells.

RSV samples were obtained and viral loads determined at multiple time points before and after the transplant and rise in lymphocyte count, thereby allowing study of longitudinally collected samples from an individual infection and under distinct levels of immunological and treatment-related selective pressure. Here, we use whole-genome deep sequencing to characterize the extent and nature of this diversity. We address the following questions. (i) Is there viral diversity in the absence of adaptive immune pressure? (ii) Is there viral diversity in the presence of adaptive immune pressure? (iii) Does the amount of diversity differ between these two states? (iv) What does the nature and context of the diversity imply about the within-host RSV dynamics and response to immune pressure?

MATERIALS AND METHODS

RSV sample preparation. Twenty-six serial nasal aspirates were obtained quantitatively and prospectively from the patient as previously described and maintained at -80°C (19). The study was conducted with the approval of the University of Tennessee Institutional Review Board and included appropriate informed consent, complying with all relevant federal guidelines and institutional policies. Viral RNA was extracted from the thawed samples using a QIA Symphony DSP virus/pathogen kit (Qiagen) per the manufacturer's protocol, including the provided synthetic poly(A). Viral RNA was eluted with 60 μl of AVE buffer and stored at -80°C . All samples were treated with Turbo DNase (Life Technologies) using the manufacturer's rigorous treatment to ensure removal of DNA.

Deep sequencing and variant calling. RNA amplification was done as previously described (20). Illumina library construction was performed using NexteraXT (Illumina) by following the manufacturer's protocol. Sequencing was performed on the Illumina HiSeq2500 platform, generating paired-end 101-bp reads. To determine the host and virus composition of each library, reads were aligned to viral and host reference genomes as previously described (20). The reference genomes used were the concatenation of the human genome assembly (GRCh37/hg19), human rRNA sequences (NR_003286.1, NR_003287.1, V00589.1, NR_003285.2, gi|251831106:648-1601, and gi|251831106:1671-3229), and the RSV-A sequence (M74568.1). The reads were assembled using the VICUNA assembly program (21). Variant calling was performed using V-Phaser 2 (22) to compare reads from each time point to the consensus sequence from the first time point. The diversity of sites was visualized using Web-Logo 3.3 (23).

Calculation of information content. For each time point, the population diversity at each amino acid position was identified using V-Phaser 2 (22). For variant positions in which the dominant amino acid was no more than 95% of the overall population, the Shannon diversity index (H_i) was determined with the following equation: $H_i = -\sum p_{a,i} \log_2 p_{a,i}$, where $p_{a,i}$ is the relative frequency of an amino acid at position i . The average diversity index for the RSV population at each time point was then determined by summing H_i over all variant positions as defined above and dividing by the length of the consensus genome sequence from the first time point, which was used as the reference for read-based mapping.

Analysis of average diversity. The average diversity indices were plotted by time. To determine whether diversity differed before and after the rise in lymphocyte count at day +12, the data were fitted to the linear regression model $Y = \beta_0 + \beta_1 A + \beta_2 (A \times I)$, where A is the explanatory variable, the β terms are the regression coefficients, and I is an indicator for the final six time points (after the rise in lymphocyte count).

Comparison of within-host and population-wide diversity. To determine population-wide diversity, RSV-A full-genome nucleotide sequences were downloaded from GenBank (see Table S1 in the supplemental material). Each gene sequence was extracted, visualized, and translated to protein sequence in SeaView (24). The Shannon diversity index was determined as described above, ignoring sites containing degenerate nucleotide calls. For the purposes of comparison of the population-wide diversity to within-host diversity, the maximum diversity at each site from all of the within-host samples was used.

Nucleotide sequence accession numbers. Sequences determined in the course of this work have been placed in GenBank under accession numbers KF973319 to KF973340.

RESULTS AND DISCUSSION

Deep sequencing of RSV samples. Of the 26 samples, 21 yielded high depth of coverage across the entire coding region of the genome, one yielded lower depth of coverage, and four had far fewer reads aligning across the genome (Table 1). Subsequent *de novo* assembly of the viral reads resulted in assemblies that covered 100%, $\sim 90\%$, and $<60\%$ of the coding region, respectively (see Fig. S1 in the supplemental material). The 22 assemblies that covered $>90\%$ of the genome were largely contiguous, with few in-

TABLE 1 Sequencing/assembly metrics for Illumina sequencing of longitudinal RSV-A samples

Sample day ^a	GenBank accession no.	RSV RNA copies/ml ^b	Total passed reads	% Passed reads aligned to reference	% Reference covered by assembly	Avg read coverage
D -35	KF973319	3.16E + 06	14,967,434	0.74	100	964.415
D -34	KF973320	2.14E + 07	11,047,458	3.75	100	3,392.926
D -33	KF973321	9.55E + 06	9,462,236	1.74	100	1,369.774
D -32	KF973322	6.03E + 06	11,177,160	3.97	100	3,707.927
D -31	KF973323	6.46E + 06	6,976,464	6.13	100	3,606.695
D -30	KF973324	1.20E + 07	8,314,566	3.53	100	2,500.276
D -29	NA ^d	3.80E + 06	10,737,830	0.05	61.5 ^c	93.47
D -27	KF973325	3.24E + 06	10,014,508	8.33	100	7,070.445
D -22	KF973326	5.89E + 06	10,845,850	1.18	100	1,097.51
D -17	KF973327	8.91E + 06	10,660,424	3.72	100	3,246.077
D -14	KF973328	6.76E + 06	11,967,902	9.97	100	9,862.39
D -10	KF973329	1.23E + 07	8,288,046	3.1	100	2,159.024
D -8	KF973330	3.16E + 06	10,717,942	0.65	100	620.338
D -2	KF973331	6.46E + 06	8,788,086	6.57	100	4,823.746
D +1	KF973332	7.41E + 06	10,493,098	0.68	100	631.46
D +4	KF973333	1.29E + 07	12,311,326	2.25	100	2,303.863
D +8	KF973334	1.62E + 07	8,718,476	20.2	100	14,563.966
D +14	NA	4.07E + 05	7,932,722	0.06	67.2 ^c	71.7
D +18	NA	7.24E + 04	27,690,936	0.01	0	NA
D +22	KF973335	1.10E + 07	9,231,440	10.64	100	8,152.972
D +26	KF973336	1.70E + 06	6,596,080	1.43	100	811.788
D +28	KF973337	6.61E + 05	8,629,562	0.18	92.6	163.338
D +32	KF973338	2.24E + 07	10,535,286	10.04	100	8,539.512
D +35	KF973339	1.02E + 07	10,179,632	0.67	100	609.665
D +39	KF973340	4.57E + 06	15,089,126	1.54	100	1,912.278
D +42	NA	1.70E + 06	9,281,832	0.03	40.56 ^c	58.01

^a Patient samples were taken at various days pre- and post-bone marrow transplantation, with the day of transplant indexed by convention as day 0.

^b RNA copies determined by RSV-A-specific quantitative reverse transcription-PCR.

^c Assemblies are fragmented with many gaps between sequence contigs.

^d NA, not applicable.

tervening gaps, while the remaining 4 assemblies contained multiple large gaps. Failure to generate contiguous assemblies covering the entire genome in these four samples was largely due to low viral loads as detected by quantitative culture (Table 1) and high levels of human nucleic acid contamination (see Table S2). These four assemblies were not used in subsequent analyses.

Reads were mapped to the *de novo* assembled consensus sequence from the first time point and amino acid variants were determined using Vicuna and V-Phaser2 software, respectively (21, 22). To assess population diversity at each time point, the genome-wide average Shannon diversity index was calculated. Fitting the plot of diversity index against time using a regression model (Fig. 1) demonstrates that there is a significant increase in the rate of accumulation of diversity after the rise in absolute CD8 cell count (ACD8) and absolute lymphocyte count (ALC) (reference 19 and Fig. 1B therein). Variant sites in which the majority population is less than or equal to 95% of the overall population are shown in Fig. 2. Analysis for insertions and deletions identifies only a single insertion with a population prevalence of >5%; this insertion, at nucleotide position 232 in the G protein, leads to a frame-shifted protein with a premature stop codon 24 amino acids downstream.

Haplotype analysis focused on regions with variants clustered by proximity along the genome; the length of the sequencing reads limited the distance over which haplotypes could be inferred. Based on this analysis, cooccurring mutations that exceed 5% of the overall number of reads covering the region include (i) the G

protein transmembrane domain (L54P and I60T), (ii) the G protein proximal ectodomain (L71P and L97P), and (iii) the N-terminal region of the M2 protein, in which proline at amino acid (aa) residue 31 was associated with several variants, including P32L, P32H, A34V, and A34T, whereas serine and glutamine at aa 31 were not associated with these variants (see Table S3 to S5 in the supplemental material).

Patterns of diversity prior to and in the presence of transplant-related immune response. The observation of diversity in the RSV population from day 35 before transplant (D -35) to D +8 (Fig. 1), prior to the rise in ALC/ACD8 and cellular immunity, could be due to neutral drift and/or to selection within the host under innate immune or other immune pressures. In addition, some or all of the diversity seen at these time points may have been present in the inoculum infecting the patient, as can be seen with other RNA viruses, such as HIV (25). Notably, most sites that appear heterogeneous prior to day +12 show no diversity subsequently (Fig. 2, G protein aa 140, 142, 150, 152, and 158). This likely reflects the impact of the RSV population bottleneck and, thereby, a reduction in genetic variation at the time of posttransplant proliferation of lymphocytes.

The increase in diversity after day +12 is attributable to an increase in the number of variant sites rather than an increase in the amount of diversity at sites that previously had small amounts of diversity (Fig. 2). Sites with markedly increased diversity after day +12 include aa 78 and 102 in NS2, P-207, M-50, M-53, and M-54; numerous sites in the transmembrane and proximal ecto-

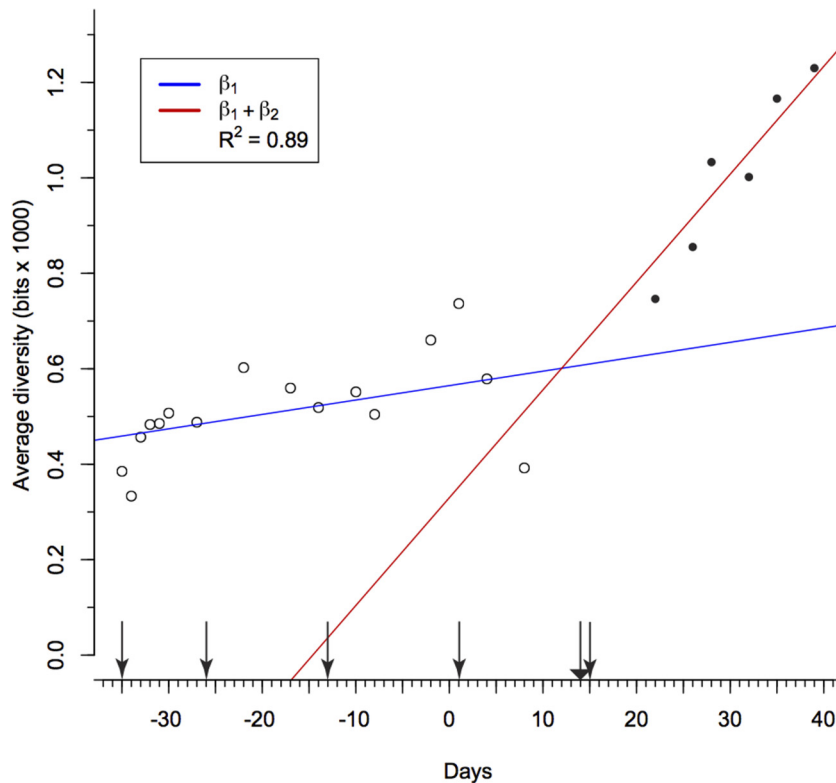


FIG 1 Plot of average diversity across the genome and lines reflecting the linear regression model coefficients; the R^2 value refers to the overall fit of the model. Note that day +12 is the transition point, as it is the date of rise in absolute lymphocyte count (ALC) and absolute CD8 cell count (ACD8) after stem cell infusion (see reference 19 and Fig. 1B therein). Open and closed circles represent diversity of samples prior to and after this point, respectively. The narrow arrowheads indicate dates of palivizumab administration, and the wide arrowhead indicates the date of administration of polyclonal RSV-Ig.

odomain of the G protein, as well as G-192 and F-268; and aa 184 and 1280 in the L protein. Despite diversity demonstrating minority variants prior to the rise of lymphocyte count, no switch in amino acid majority predominance occurs. After day +12, the predominant amino acids switch at multiple sites (e.g., NS2-102, P-207, G-56, G-60, G-292, F-268, and L-1280) (Fig. 2).

Several amino acid sites, including M2-31, G-229, and G-292, show diversity that persists across all time points. In the case of G-229, one subpopulation contains a premature stop codon, truncating a large segment of the carboxy (C) terminus. The functional implications of the other variants are unclear, but the persistent polymorphisms at all three sites have at least two possible causes: balancing selection driven by a pressure other than adaptive immunity, or sampling of populations from more than one distinct subpopulation that do not directly compete with one another, thereby maintaining polymorphisms.

Diversity in the G protein. The greatest diversity observed in the samples appears in the G protein, which contains 45% of the variant amino acid sites across all samples. The diversity in the G protein is consistent with its role as a major surface antigen of RSV and relaxed selective constraints. The G protein product has two forms: a transmembrane form (G_m), which functions as the attachment glycoprotein for RSV, and a secreted form (G_s), which is cleaved just downstream from the transmembrane domain and is believed to function as an antigen decoy to allow evasion of antibody targeting (26, 27) and as a modulator of dendritic cell function (28).

The diversity within the G protein is notable for several features. First, variants in G protein appear concentrated in three regions (Fig. 3). These are (i) aa 33 to 97, extending across the transmembrane domain and into the subsequent proximal ectodomain; (ii) aa 135 to 166, which includes the mucin-like domain and extends slightly into the central conserved domain; and (iii) aa 204 to 292, which includes the mucin-like domain downstream from the central conserved domain. Second, the diversity in these regions shows patterns that appear to change based on the presence or absence of host functional lymphocytes. Diversity in variable region ii is seen predominantly before the rise in ACD8/ALC (Fig. 2). Diversity in variable region iii appears across all time points, and diversity in variable region i is seen predominantly after the rise in ACD8/ALC. Third, the within-host diversity seen in the G protein contains features that have not been reported in epidemiological studies of RSV-A (discussed below), raising questions about the extent to which consensus sequences mask true viral population diversity and the role of G protein variants in RSV quasi-species fitness in transmission and infection.

The within-host diversity of the transmembrane domain is unexpected, as the transmembrane domain is frequently highly conserved in molecular epidemiology studies of the RSV-A G protein (Fig. 3) (13). However, escape mutations in the transmembrane domain have been demonstrated to emerge under selection by a neutralizing antibody targeting the G protein (10). In that example, reduced amounts of full-length G protein were detected on the cell surface and incorporated into the virion, and increased

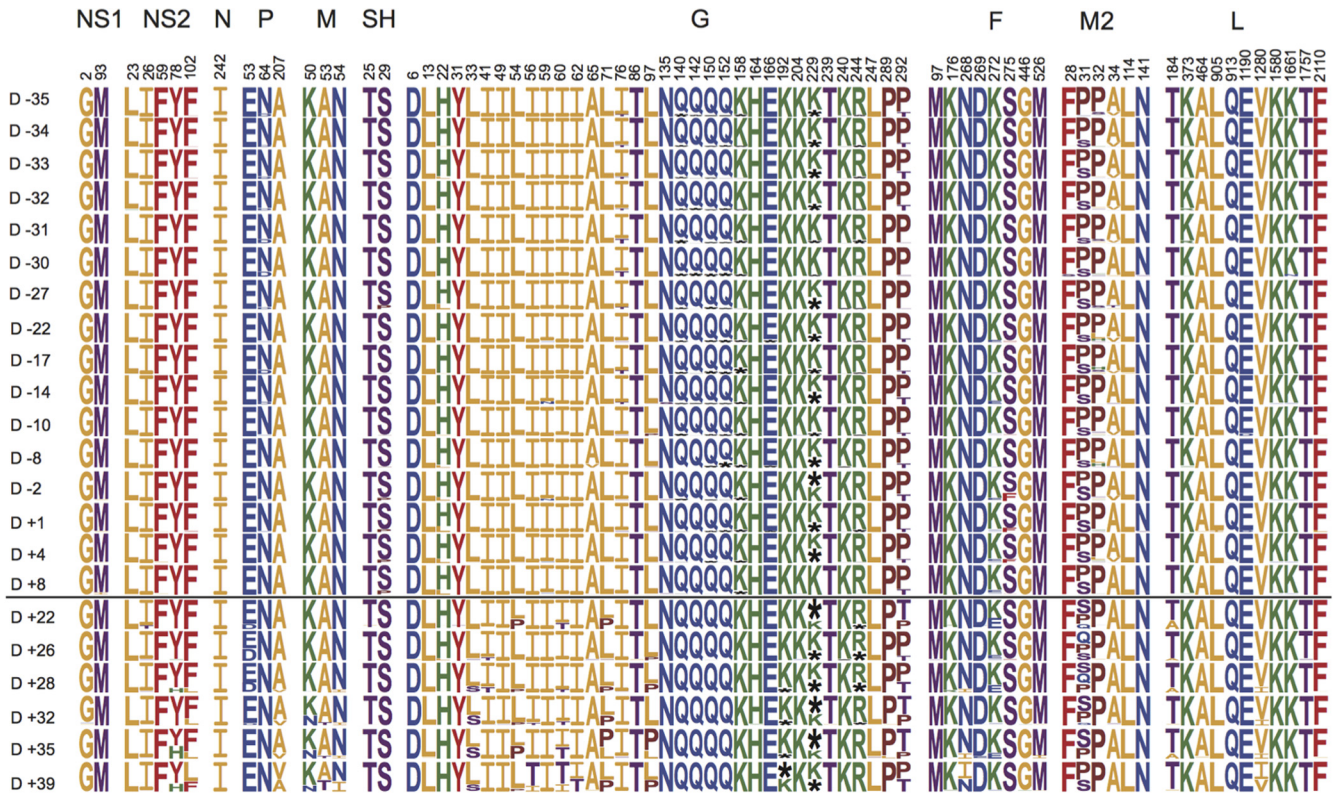


FIG 2 Variable amino acids in the RSV samples. Each sample date is represented by one row. Genes are indicated in the top row, and amino acid sites (indexed to the consensus sequence determined for the genome from the sample on day -35 [D -35]) below. Amino acids are considered variable if the dominant amino acid call is at most 95% of the overall population for at least one of the time points. The height of the amino acid's designated letter reflects its proportion in the overall population. Asterisks represent stop codons. The horizontal line between days $+8$ and $+22$ represents the time after which lymphocytes and CD8 cells became present (day $+12$).

amounts of the secreted form were seen in the cell supernatant. These results could reflect that mutations in the transmembrane domain lead to diminished incorporation of the G protein into the virion, or that the mutations lead to greater susceptibility to cleavage resulting in Gs. The observation, then, of transmembrane domain variants indicates an adaptive response to the pressure of posttransplant immunity, either through diminished levels of Gm leading to decreased antibody binding or elevated levels of Gs functioning as decoys and decreasing the amount of available soluble antibody.

Similarly, while a number of truncated G proteins have been observed in RSV-B *in vivo* (29), we are unaware of reports of truncated G proteins in RSV-A *in vivo*. Studies of RSV escape from antibody-mediated selection have shown that premature stop codons in the C-terminal region can lead to antibody escape in cell culture (11). The presence of premature stop codons at every time point, however, suggests that the emergence of C-terminally truncated G protein is not due solely to immune pressure. In addition, the presence of minority populations with premature stop codons in the first mucin-like domain, prior to the heparin binding domain, during the pretransplant time points raises questions about whether the variant-containing viruses are viable or fit. While *in vitro* and mouse *in vivo* experiments have demonstrated that RSV lacking the G protein can still infect cells, on the whole such virus appears to infect less efficiently and replicate more slowly (30, 31).

Diversity in the F protein. In the initial study of these samples,

consensus sequencing of the palivizumab binding domain of the fusion (F) protein (aa 186 to 315) revealed no changes across the sample set despite multiple administrations of palivizumab (19). In contrast, deep sequencing of the viral populations (Fig. 2) shows that minority populations contain known palivizumab escape mutations at aa 268 (N268I [32]), 272 (both K272E and K272N [33, 34]), and 275 (S275F [33, 34]); haplotype analysis demonstrates that the variant alleles at positions 268, 272, and 275 occurred in separate populations. The presence of these minority populations only after day -22 (13 days after the start of palivizumab infusions) suggests that palivizumab exerted a selective pressure in this population, consistent with what has been observed rarely in palivizumab-treated populations (35).

The N268I variant at day $+39$ is the only F protein variant that becomes dominant. It is notable that the N268I variant does not appear until day $+28$, after multiple administrations of palivizumab, single-dose polyclonal RSV-Ig, and lymphocyte infusion at the time of transplant. The more commonly epidemiologically observed palivizumab escape variants at aa 272 and 275 appear during the pre-engraftment period but remain minority alleles despite multiple rounds of palivizumab administration, likely reflecting levels of palivizumab insufficient to generate enough selective pressure on the viral population to overcome the presumed fitness cost of the variants. It is also notable that the clinical samples evaluated all were obtained from the upper respiratory tract (nasal aspirates), and the concentrations of functional RSV-neu-

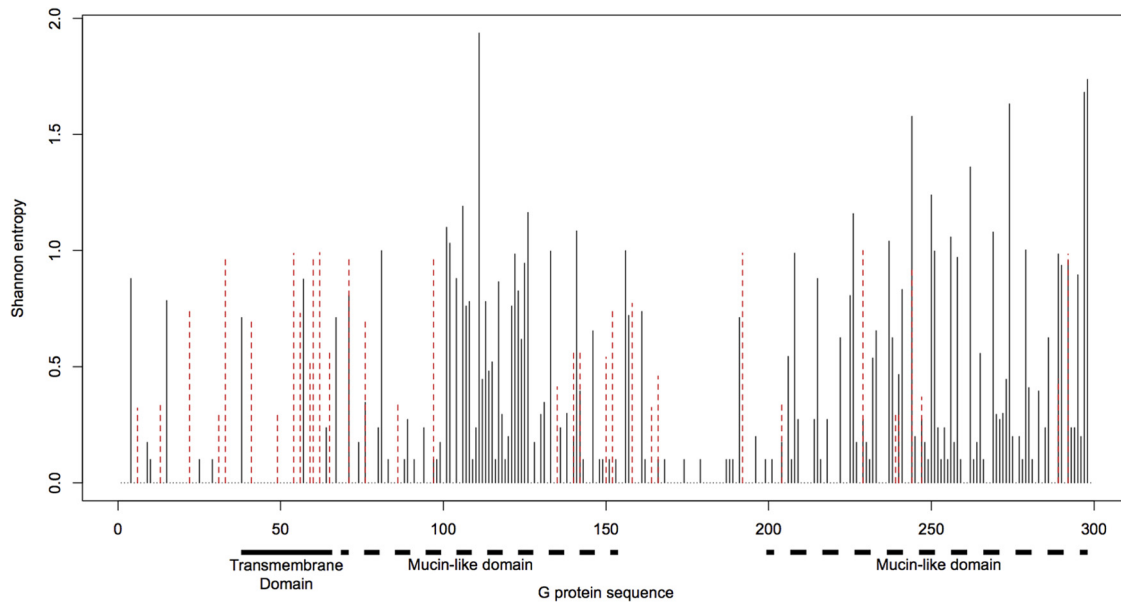


FIG 3 Comparison of within-host and population-level G protein diversity. Black lines indicate the Shannon diversity index at each residue in the population-wide analysis, and dashed red lines indicate the maximum Shannon diversity index across all time points in the within-host samples. See “Comparison of within-host and population-wide diversity” in Materials and Methods for a description of the analytical approach using the sequences reported in Table S1 in the supplemental material.

tralizing palivizumab achievable in this anatomic site are approximately 10-fold lower than those in the deeper lower respiratory tract (36). Thus, palivizumab immune pressure measured in the upper tract may not have been great enough to create and sustain a consensus amino acid mutation within this anatomic site. The additional mutations observed at aa 97, 176, 269, 446, and 526 predominantly appear after the rise in ACD8 and may reflect the influence of specific or nonspecific immune responses. The impact of these additional mutations on F protein function and viral fitness are uncertain but bear further investigation as functional immune epitopes.

Diversity in other regions of the genome. Variant sites observed in the other 9 genes in the genome that arise after bone marrow transplant and subsequent increase in ACD8/ALC may reflect sites undergoing immune selection. Notable sites include aa 78 and 102 in the nonstructural protein 2 (NS2; thought to modulate the immune response through interaction with RIG-I and decreased TRAF3 and STAT2); aa 53 and 207 in the phosphoprotein (P protein), thought to be a key component of the viral RNA-dependent RNA polymerase complex); aa 50, 53, and 54 of the matrix (M) protein; and aa 184 and 1280 in the polymerase (L protein). Whether these variants emerged due to selection from specific or nonspecific immunity, genetic hitchhiking (when the frequency of a variant changes due to linkage with a positively associated variant elsewhere in the genome), or neutral drift is unclear. Comparison of the patterns of diversity in these genes in the within-host sample and the samples from molecular epidemiological studies shows no clear overlap (see Fig. S2A to I in the supplemental material).

While the rate of nonsynonymous versus synonymous substitutions (dN/dS) can be helpful in specific situations in determining evidence of positive selection, dN/dS is difficult to interpret for deep sequencing data of a population in which variants have not

yet fixed (37). Interpretability is further undermined by not being able to ascribe linkage across full gene sequences: synonymous sites might be polymorphic due to hitchhiking with nonsynonymous sites in a way that cannot be ascertained due to the short sequence read lengths. Moreover, even without this problem, in general, positive selection need not lead to elevated values of dN/dS for within-population polymorphisms (37). Thus, in the absence of a clear theoretical basis for interpreting the value of the dN/dS statistic in the context of these deep sequencing data, we do not present it and agree with the need for further study and development of methods that permit interpretation of evidence of evolutionary pressure in similar deep sequencing data sets.

Conclusions. This study presents the first whole-genome deep sequencing analysis of clinical RSV samples, and the specific aspects of this chosen clinical case offer a unique opportunity to evaluate the impact of differing levels of immune function on the nature and extent of within-host viral population diversity. We observe baseline diversity in the initial set of samples prior to bone marrow transplant and increasing diversity correlating with the presence of donor lymphocytes. The initial diversity may reflect a mixed inoculum, drift, or anatomically distinct populations that are captured in the upper respiratory clinical sampling procedure. The data demonstrate increasing diversity in the presence of adaptive immunity, though it is unclear to what degree the diversity is due to specific versus nonspecific inflammatory processes.

Through the detailed description of pathogen variation over the course of an individual infection, this study raises a number of questions regarding RSV pathogenesis and evolution. Do the patterns of variation observed in this example represent the unique circumstance of longstanding RSV infection and the combination of bone marrow transplant and immunosuppression, or are they representative of general RSV dynamics during individual infections? Do the variable sites which emerge after the rise in lympho-

cyte count reflect specific or nonspecific immune pressure, other selective pressures, or neutral drift? Does the abundance of variants in the G protein transmembrane domain indicate greater production of Gs as an antibody decoy or another process, and does the rare appearance of transmembrane domain variants in population-wide surveys of G protein reflect the limited ability of these variants to transmit and cause new infections? Are the variable sites seen in genes coding for nonsurface antigens evidence of selection in these genes, and if so, through what mechanism? Additional deep sequencing whole-genome studies over the course of individual infections in immune-naïve and immune-experienced immunocompetent hosts will help address many of these questions.

ACKNOWLEDGMENTS

This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. HHSN272200900018C. M.L. was supported by grant U54GM088558 from the National Institute of General Medical Sciences. Internal funds from the DeVincenzo laboratory were also used for sample processing and technique development.

We thank Tiffany Poon and Jennifer Kelliher for their technical assistance.

REFERENCES

- Hall CB. 2010. Respiratory syncytial virus, p 2207–2221. *In* Mandell GL, Bennett JE, Dolin R (ed), Principles and practice of infectious diseases, vol 2. Churchill Livingstone Elsevier, Philadelphia, PA.
- Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. 2005. Respiratory syncytial virus infection in elderly and high-risk adults. *N. Engl. J. Med.* 352:1749–1759. <http://dx.doi.org/10.1056/NEJMoa043951>.
- Kim YJ, Boeckh M, Englund JA. 2007. Community respiratory virus infections in immunocompromised patients: hematopoietic stem cell and solid organ transplant recipients, and individuals with human immunodeficiency virus infection. *Semin. Respir. Crit. Care Med.* 28:222–242. <http://dx.doi.org/10.1055/s-2007-976494>.
- Lujan-Zilbermann J, Benaim E, Tong X, Srivastava DK, Patrick CC, DeVincenzo JP. 2001. Respiratory virus infections in pediatric hematopoietic stem cell transplantation. *Clin. Infect. Dis.* 33:962–968. <http://dx.doi.org/10.1086/322628>.
- Collins PL, Melero JA. 2011. Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years. *Virus Res.* 162: 80–99. <http://dx.doi.org/10.1016/j.virusres.2011.09.020>.
- Glezen WP, Taber LH, Frank AL, Kasel JA. 1986. Risk of primary infection and reinfection with respiratory syncytial virus. *Am. J. Dis. Child.* 140:543–546.
- Hall CB, Walsh EE, Schnabel KC, Long CE, McConnochie KM, Hildreth SW, Anderson LJ. 1990. 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.* 162: 1283–1290. <http://dx.doi.org/10.1093/infdis/162.6.1283>.
- Peret TC, Hall CB, Hammond GW, Piedra PA, Storch GA, Sullender WM, Tsou C, Anderson LJ. 2000. Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. *J. Infect. Dis.* 181:1891–1896. <http://dx.doi.org/10.1086/315508>.
- Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ. 1998. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. *J. Gen. Virol.* 79(Part 9):2221–2229.
- Walsh EE, Falsey AR, Sullender WM. 1998. Monoclonal antibody neutralization escape mutants of respiratory syncytial virus with unique alterations in the attachment (G) protein. *J. Gen. Virol.* 79(Part 3):479–487.
- Rueda P, Delgado T, Portela A, Melero JA, Garcia-Barreno B. 1991. Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies. *J. Virol.* 65:3374–3378.
- Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, Vieira SE, Stewien KE, Peret TC, Jamal LF, Pardini MI, Pinho JR, Massad E, Sant'anna OA, Holmes EC, Durigon EL. 2009. Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. *PLoS Pathog.* 5:e1000254. <http://dx.doi.org/10.1371/journal.ppat.1000254>.
- Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, van Loon AM, Wiertz E, van Bleek GM, Martin DP, Coenjaerts FE. 2012. Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology. *PLoS One* 7:e51439. <http://dx.doi.org/10.1371/journal.pone.0051439>.
- Zlateva KT, Lemey P, Vandamme AM, Van Ranst M. 2004. Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup A: positively selected sites in the attachment G glycoprotein. *J. Virol.* 78:4675–4683. <http://dx.doi.org/10.1128/JVI.78.9.4675-4683.2004>.
- Oshansky CM, Zhang W, Moore E, Tripp RA. 2009. The host response and molecular pathogenesis associated with respiratory syncytial virus infection. *Future Microbiol.* 4:279–297. <http://dx.doi.org/10.2217/fmb.09.1>.
- Henn MR, Boutwell C, Lennon N, Power K, Malboeuf C, Charlebois P, Gladden A, Levin J, Casali M, Philips L, Berlin A, Berical A, Erlich R, Anderson S, Streeck H, Kemper M, Ryan E, Wang Y, Green L, Axten K, Brumme Z, Brumme C, Russ C, Rosenberg E, Jessen H, Altfeld M, Nusbaum C, Walker B, Birren B, Allen TM. 2009. P09–20 LB. Ultra-deep sequencing of full-length HIV-1 genomes identifies rapid viral evolution during acute infection. *Retrovirology* 6:P400. <http://dx.doi.org/10.1186/1742-4690-6-S3-P400>.
- Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR. 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One* 4:e5683. <http://dx.doi.org/10.1371/journal.pone.0005683>.
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303:327–332. <http://dx.doi.org/10.1126/science.1090727>.
- El Saleeby CM, Suzich J, Conley ME, DeVincenzo JP. 2004. Quantitative effects of palivizumab and donor-derived T cells on chronic respiratory syncytial virus infection, lung disease, and fusion glycoprotein amino acid sequences in a patient before and after bone marrow transplantation. *Clin. Infect. Dis.* 39:e17–20. <http://dx.doi.org/10.1086/421779>.
- Malboeuf CM, Yang X, Charlebois P, Qu J, Berlin AM, Casali M, Pesko KN, Boutwell CL, DeVincenzo JP, Ebel GD, Allen TM, Zody MC, Henn MR, Levin JZ. 2013. Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res.* 41:e13. <http://dx.doi.org/10.1093/nar/gks794>.
- Yang X, Charlebois P, Gnerre S, Coole MG, Lennon NJ, Levin JZ, Qu J, Ryan EM, Zody MC, Henn MR. 2012. De novo assembly of highly diverse viral populations. *BMC Genomics* 13:475. <http://dx.doi.org/10.1186/1471-2164-13-475>.
- Yang X, Charlebois P, Macalalad A, Henn MR, Zody MC. 2013. V-Phaser 2: variant inference for viral populations. *BMC Genomics* 14:674. <http://dx.doi.org/10.1186/1471-2164-14-674>.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190. <http://dx.doi.org/10.1101/gr.849004>.
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–224. <http://dx.doi.org/10.1093/molbev/msp259>.
- Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Tomaras WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* 105:7552–7557. <http://dx.doi.org/10.1073/pnas.0802203105>.
- Bukreyev A, Yang L, Fricke J, Cheng L, Ward JM, Murphy BR, Collins PL. 2008. The secreted form of respiratory syncytial virus G glycoprotein helps the virus evade antibody-mediated restriction of replication by act-

- ing as an antigen decoy and through effects on Fc receptor-bearing leukocytes. *J. Virol.* 82:12191–12204. <http://dx.doi.org/10.1128/JVI.01604-08>.
27. Melero JA, Garcia-Barreno B, Martinez I, Pringle CR, Cane PA. 1997. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. *J. Gen. Virol.* 78(Part 10):2411–2418.
 28. Shingai M, Azuma M, Ebihara T, Sasai M, Funami K, Ayata M, Ogura H, Tsutsumi H, Matsumoto M, Seya T. 2008. Soluble G protein of respiratory syncytial virus inhibits Toll-like receptor 3/4-mediated IFN- β induction. *Int. Immunol.* 20:1169–1180. <http://dx.doi.org/10.1093/intimm/dxn074>.
 29. Venter M, van Niekerk S, Rakgantso A, Bent N. 2011. Identification of deletion mutant respiratory syncytial virus strains lacking most of the G protein in immunocompromised children with pneumonia in South Africa. *J. Virol.* 85:8453–8457. <http://dx.doi.org/10.1128/JVI.00674-11>.
 30. Techaarpornkul S, Barretto N, Peeples ME. 2001. Functional analysis of recombinant respiratory syncytial virus deletion mutants lacking the small hydrophobic and/or attachment glycoprotein gene. *J. Virol.* 75:6825–6834. <http://dx.doi.org/10.1128/JVI.75.15.6825-6834.2001>.
 31. Teng MN, Whitehead SS, Collins PL. 2001. Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication in vitro and in vivo. *Virology* 289:283–296. <http://dx.doi.org/10.1006/viro.2001.1138>.
 32. Zhao X, Chen FP, Megaw AG, Sullender WM. 2004. Variable resistance to palivizumab in cotton rats by respiratory syncytial virus mutants. *J. Infect. Dis.* 190:1941–1946. <http://dx.doi.org/10.1086/425515>.
 33. Zhu Q, McAuliffe JM, Patel NK, Palmer-Hill FJ, Yang CF, Liang B, Su L, Zhu W, Wachter L, Wilson S, MacGill RS, Krishnan S, McCarthy MP, Losonsky GA, Suzich JA. 2011. Analysis of respiratory syncytial virus preclinical and clinical variants resistant to neutralization by monoclonal antibodies palivizumab and/or motavizumab. *J. Infect. Dis.* 203:674–682. <http://dx.doi.org/10.1093/infdis/jiq100>.
 34. Zhu Q, Patel NK, McAuliffe JM, Zhu W, Wachter L, McCarthy MP, Suzich JA. 2012. Natural polymorphisms and resistance-associated mutations in the fusion protein of respiratory syncytial virus (RSV): effects on RSV susceptibility to palivizumab. *J. Infect. Dis.* 205:635–638. <http://dx.doi.org/10.1093/infdis/jir790>.
 35. Papenburg J, Carbonneau J, Hamelin ME, Isabel S, Bouhy X, Ohoumanne N, Dery P, Paes BA, Corbeil J, Bergeron MG, De Serres G, Boivin G. 2012. Molecular evolution of respiratory syncytial virus fusion gene, Canada, 2006–2010. *Emerg. Infect. Dis.* 18:120–124. <http://dx.doi.org/10.3201/eid1801.110515>.
 36. Wu H, Pfarr DS, Johnson S, Brewah YA, Woods RM, Patel NK, White WI, Young JF, Kiener PA. 2007. Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *J. Mol. Biol.* 368:652–665. <http://dx.doi.org/10.1016/j.jmb.2007.02.024>.
 37. Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS . *PLoS Genet.* 4:e1000304. <http://dx.doi.org/10.1371/journal.pgen.1000304>.