Sequencing and Analyses of All Known Human Rhinovirus Genomes Reveal Structure and Evolution

Ann C. Palmenberg,1* David Spiro,2* Ryan Kuzmickas,2 Shiliang Wang,2 Appolinaire Djikeng,2 Jennifer A. Rathe,3 Claire M. Fraser-Liggett,4 Stephen B. Liggett†*

Infection by human rhinovirus (HRV) is a major cause of upper and lower respiratory tract disease worldwide and displays considerable phenotypic variation. We examined diversity by completing the genome sequences for all known serotypes \((n = 99)\). Superimposition of capsid structure and optimal-energy RNA configurations established alignments and phylogeny. These revealed conserved motifs; clade-specific diversity, including a potential newly identified species (HRV-D); mutations in field isolates; and recombination. In analogy with poliovirus, a hypervariable 5′ untranslated region tract may affect virulence. A configuration consistent with nonscanning internal ribosome entry was found in all HRVs and may account for rapid translation. The data density from complete sequences of the reference HRVs provided high resolution for this degree of modeling and serves as a platform for full genome-based epidemiologic studies and antiviral or vaccine development.

Human rhinovirus (HRV), the disease agent for the common cold, is responsible for ~50% of asthma exacerbations and is one of the factors that can direct the infant immune system toward an asthmatic phenotype \((I–4)\). Direct and indirect costs from the common cold and related complications in asthma and chronic obstructive pulmonary disease amount to an estimated ~$60 billion per year in the United States \((5, 6)\). HRVs are single-stranded, positive-sense RNA enteroviruses in the Picornaviridae family and have been cataloged primarily by capsid serotyping relative to a historical repository of 99 strains, obtained from clinical specimens. HRVs are classified by their use of viral and cellular proteins for the initiation of RNA transcription for certain traits, yet still have mutational flexibility for escape from immune responses.

Picornaviruses encode a single open reading frame (ORF) representing about 90% of the RNA length. Translation produces a polyprotein (~215 kD for HRV), subsequently cleaved in a viral protease–dependent cascade to form the 11 to 12 mature viral proteins required to initiate and sustain an infection \((fig. S2A)\). Local and global RNA structures play established roles in HRV biology; however, the extent, character, and relatedness of serotype-specific 5′- and 3′UTR (untranslated region) variation are unknown. The role of this variability has thus not been related to function by modeling techniques or in vivo approaches. Our alignment methods included optimal energy RNA structure considerations \((I5)\) and were therefore sensitive to potential differences among the HRV's in the 5′- and 3′UTRs.

HRV RNA structures. All enteroviruses encode 5′-terminal cloverleaf-like motifs (CLs) that bind viral and cellular proteins for the initiation of RNA synthesis and also help convert infecting genomes from translation to replication templates. The HRV CLs \([80 to 84 bases (b)]\) were predicted in every sequence with minimal structural variation among the species \((fig. S1A and additional structures in fig. S3)\). Immediately 3′ to the CL, all HRVs were found to share an unusual pyrimidine-rich spacer segment with short oligo(C) and oligo(U) units interspersed with As \((blue boxes, fig. 1A and fig. S3A)\). The HRV-A have the shortest tracts \((11 to 22 b)\) and HRV-B the longest \((22 to 50 b)\). Nearly every HRV displayed a unique sequence in this region, and we identified an unexpected variation even among isolates of the same serotype \((fig. 1A and fig. S3A)\). The equivalent genome location in poliovirus \((10 b)\) interacts with poly(C)-binding protein 2 and is involved in the determination of the polio neuroviral potential. The analogous regions in aphthoviruses or cardioviruses have homopolymeric poly(C) or poly(U) tracts, the deletion of which markedly attenuates the virus through an RNA-activated protein kinase activation–dependent mechanism \((16)\). If the HRV 5′ spacer tracts are functional analogs to those of these other picornaviruses, then it is possible that the pathogenic potential of an individual HRV may also be encoded, in part, by this region.

Picornaviruses use internal ribosome entry sites (IRESs) to mediate translation initiation of their
polypeptide ORFs. The IRESs of all enteroviruses (termed type 1 IRESs) are thought to bind 40S ribosomal subunits internally within their 5′UTR and to then scan additional nucleotides to find the proper initiator AUG (17). Our modeling of the known serotypes confirms that all HRV IRESs start just 3′ to the pyrimidine-rich spacer tract. We found that the internal IRES sequences are highly conserved, with an average nucleotide identity of 82%. Indeed, this region of the genome has the greatest degree of identity among all HRV (fig. S2B), exceeding 95% for regional motifs within the IRES. However, we also observed that the dominant IRES sequence conservation did not extend completely to the initiator AUG and that for the 18 to 40 bases 5′ to this codon, the region scanned by ribosomes, there was little species-specific conservation (<60% nucleotide identity). Despite this, our folding predictions configured every one of these regions into virtually the same RNA motif, which suggests that this structure is conserved even when the underlying nucleotides are not (Fig. 1B and fig. S3B).

Near the bottom of a long [15 to 20 base pair (bp)] minimum energy unbranched stem, the ORF AUG was invariably paired with a conserved upstream noncoding AUG (green boxes, Fig. 1B), marking the 3′ boundary of the IRES, and the normal launch point for 40S scanning. Every HRV genome fold maintains this pairing. We predict that HRVs use the proximity of these AUGs to orient the 40S for direct transfer to the proper codon, without the need for scanning through the intervening nucleotides. The sequence and length variation between the AUGs was consistent with the idea that ribosomes would bypass this region entirely if they jump from one AUG to the other. This IRES folding is essentially a bait-and-switch mechanism, which we predict may enhance HRV translation competitiveness. This paired AUG motif is unique to HRV and is not found in other enteroviruses (e.g., poliovirus).

The HRV 3′UTRs (40 to 60 b) begin with the ORF termination codon and extend to the genetically encoded poly(A) tail. The ORF terminators themselves (solid red boxes in Fig. 1C and fig. S3D) included UAG, UAA, and UGA codons. The codon selection often differed among isolates from the same serotype. Multiple additional terminators (tan boxes), in and out of the ORF, were identified that punctuated each segment (3 to 9 per UTR). Despite large differences (>40%) in nucleotide identity (Fig. 1C), it was noted that all HRVs maintain a 13- to 16-bp unbranched stem, covering 67 to 88% of the 3′UTR, immediately abutting the poly(A) tail (see also fig. S3D). Some 3′ stems have small interior loops, but nearly all, except for the unusual sequences of clade-D, present 5-base terminal loops, anchored with apical U-G or U-A pairs. Inevitably, the 3′ sides of these terminal loops display UAG or UGA terminator codons, which may or may not synchronize with the local ORF. The function of these 3′ stems is unknown, but such conservation is usually indicative of a putative protein recognition motif (such as translation termina-

Phylogenetic relationships of the HRV. Multiple methods were used to compute and compare phylogenetic trees for the aligned RNA and protein data. Figure 2 is a neighbor-joining consensus tree (18), which considered the 5′- and 3′UTRs and the first and second codon positions for the RNA genomes. All major nodes of the tree topology were stable over a range of calculation parameters, regardless of whether they invoked minimum evolution, parsimony, unweighted pair group method with arithmetic mean, or maximum likelihood (ML) methods. (See fig. S6A for ML tree with bootstrap, likelihood ratio tests, and comparison to the neighbor-joining tree.) Statistical comparisons were performed between the ML tree generated from the full genomes and the ML constrained-topology trees generated from capsid sequences that have been used as a surrogate for serotypes (fig. S6, A to C) (13), with the null hypothesis that topologies of these limited-sequence trees were the same as that of the full genome-based tree. The approximately unbiased (AU) test calculated from the multiscale bootstrap P values for the capsid VP1 and VP0 trees were <10−69, strongly rejecting the null hypothesis. In a similar manner, ML constrained-topology trees generated from IRES and 3D-polymerase sequences were not consistent with the full genome tree (fig. S6, A and B). Additional statistical tests also confirmed these results from the AU (fig. S6, B and C) for all four of the constrained-topology trees. Taken together, these results suggest that trees based on sequences from small, albeit biologically important, regions of the HRV genome have a limited capacity to reflect the evolutionary relationships and underlying diversity of HRVs. The tree shown in Fig. 2 also accurately represents parallel calculations on the aligned polyprotein dataset (p-distances differed by <2%). All aligned full-length sequences contributed to the tree calculations, but published sequences (noted with asterisks on Fig. 2) are shown only as needed, to complete the cohort of reference serotypes. (Were the redundant published sequences illustrated, without exception they would lie on the same serotype branches, with p-distances <0.5%) The outer rings of this figure designate major (M: ICAM-1) or minor (m: LDLR) receptor preferences, and also each

Fig. 1. Genome-wide optimal energy RNA configurations for select motifs representative of each HRV species. (A) All HRVs display characteristic 5′UTR CL elements with minimal predicted structural variation (blue boxes). (B) The alignments and predicted structures of the IRES of HRVs reveal a bait-and-switch arrangement for initiation of translation. A stem-pairing AUG (light green box) is found for each species, but a highly variable region of intervening sequence near the initiator AUG (dark green box) was noted. (C) HRV 3′UTRs have a unique unbranched stem motif before the poly(A) tail. The left-most codon (red box, white text) is the ORF terminator. Other boxes highlight additional terminators (tan boxes), including a characteristic codon (tan box, red text) found in the apical loop.
The HRV-C have yet to be cultured or assessed for immunological cross-reactivity, but the sequence space occupied by the available samples suggests that there may be many additional HRV-C strains awaiting discovery. Distance extrapolations relative to the new full reference cohort predict that HRV-C may have an even broader range of serotypes than the original 99, of which each confers only limited immunologic cross-protection to another.

A separate phylogenetic finding was the unexpected basal divergence within HRV-A of a small \( (n = 3) \) group of distinct strains, denoted clade D \( (20) \). Although the major basis for discriminating clade D from other HRV-A lies in their general, genome-length sequence divergence, these particular isolates have RNA elements—such as the cis-acting replication element, the 3′UTR terminal loop feature (see above and fig. S3), and local insertions/deletions and sequence motifs—that are somewhat atypical of other HRV-A strains. Some of the distinguishing characteristics are highlighted in fig. S4. Among all other major A clades, and major B clades, none have p-distances (>10%) that segregate them so distinctly. We are cautious in proposing clade D as a fourth species (HRV-D), but the phylogenetic evidence (fig. 2) and sequence characteristics (figs. S3 and S4) are highly suggestive. Other early topological divisions within HRV-A separate a major clade composed of 10 serotypes (counterclockwise, hrv-20 through hrv-12) from a second grouping composed of ~12 miniclaides representing 61 serotypes (counterclockwise, hrv-89-f09 through hrv-100). These particular relationships were not readily apparent when only partial genome sequences were examined \( (21, 22) \). Fig. S6 shows a comparison of results from trees constructed with whole genomes, VP4/VP2, and VP1 sequence. Neither of the trees derived from these shorter sequences revealed the miniclaides, clade D, or multiple other features. We thus contend that comparison of full-genome data, the context wherein evolutionary events occur, most likely provides the defining relationships among the HRVs, allows a more comprehensive assessment of strain diversity, and allows for more accurate historical extrapolations. The phylogenetic diversity we describe at the whole-genome level is consistent with the clinical heterogeneity of HRV infections in humans \( (1, 3, 4) \), although mapping specific clinical characteristics (i.e., incubation period, severity, respiratory compromise, and pro-asthmatic phenotypes) to responsible genomic regions will require additional field isolates from a large number of patients with multiple traits. Given the genome-wide diversity we have documented (e.g., 5′ spacer elements, ORF start, protease, 3′UTRs), clinically relevant relationships may well depend on comparisons from multiple genome regions.

**Recombination in HRV.** Results from earlier sequencing of a subset of HRV reference genomes concluded that RNA recombination was not a major mechanism for HRV diversity \( (23, 24) \) and asserted that known isolates were independently segregating entities. We have reevaluated the potential for recombination by scanning the full reference set and the new field strains with a suite of recombination detection programs \( (25) \) relying on phylogenetic distance and sequence similarity. Stringent criteria \( (P < 0.00001\) from two or more analyses modes) identified 23 genomes with probable origins resulting from at least 12 independent recombination events. Figure 3A shows representative data indicating that hrv-46 arose by recombination between hrv-53 (major parent) and hrv-80 (minor parent). Within the hrv-46 genome, nucleotides 32 to 3222 are most similar to hrv-80, whereas the rest of the genome (nucleotides 3223 to 7200) is common to hrv-53. The result is consistent with this trio’s computed phylogenetic relationship (fig. 2), placing the major parent (hrv-53) and the daughter (hrv-46) in the same clade and the minor parent (hrv-80) in a different, nearby clade. Results for all 23 identified recombination scenarios are summarized in Table 1. [See also \( (13) \) and table S4.] Of the recombination locales suggested by these events, the majority (10 of 12) involve the 5′UTR or the adjacent capsid genes, which seemingly have been collectively rearranged to produce at least 20 separate progeny strains. Among the
Table 1. Recombination events in HRV serotypes. The P value listed is the lowest obtained. See table S4 for full data and P values.

<table>
<thead>
<tr>
<th>Major parent</th>
<th>Minor parent</th>
<th>Recombinant</th>
<th>Genome region</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrv-45</td>
<td>hrv-21</td>
<td>hrv-8</td>
<td>5’UTR</td>
<td>1.094 × 10⁻²¹</td>
</tr>
<tr>
<td>hrv-45</td>
<td>hrv-21</td>
<td>hrv-95</td>
<td>5’UTR</td>
<td>1.094 × 10⁻²¹</td>
</tr>
<tr>
<td>hrv-65</td>
<td>hrv-21</td>
<td>hrv-80</td>
<td>5’UTR, VP4</td>
<td>6.129 × 10⁻²³</td>
</tr>
<tr>
<td>hrv-51</td>
<td>hrv-11</td>
<td>hrv-20</td>
<td>5’UTR, VP4</td>
<td>8.402 × 10⁻²⁶</td>
</tr>
<tr>
<td>hrv-51</td>
<td>hrv-11</td>
<td>hrv-68</td>
<td>5’UTR, VP4</td>
<td>8.402 × 10⁻²⁶</td>
</tr>
<tr>
<td>hrv-28</td>
<td>hrv-62</td>
<td>hrv-71</td>
<td>5’UTR</td>
<td>8.484 × 10⁻¹⁵</td>
</tr>
<tr>
<td>hrv-54-405</td>
<td>hrv-75</td>
<td>hrv-18</td>
<td>5’UTR, VP4, VP2, VP3, VP1</td>
<td>3.737 × 10⁻⁷</td>
</tr>
<tr>
<td>hrv-54-405</td>
<td>hrv-75</td>
<td>hrv-24</td>
<td>5’UTR, VP4, VP2, VP3, VP1</td>
<td>3.737 × 10⁻⁷</td>
</tr>
<tr>
<td>hrv-54-405</td>
<td>hrv-75</td>
<td>hrv-50</td>
<td>5’UTR, VP4, VP2, VP3, VP1, P2A</td>
<td>3.737 × 10⁻⁷</td>
</tr>
<tr>
<td>hrv-54-405</td>
<td>hrv-75</td>
<td>hrv-34</td>
<td>VP4, VP2, VP3, VP1</td>
<td>3.737 × 10⁻⁷</td>
</tr>
<tr>
<td>hrv-54-405</td>
<td>hrv-67</td>
<td>hrv-38</td>
<td>VP2, VP3, VP1, P2A, P2B, P2C, P3A</td>
<td>1.561 × 10⁻²⁸</td>
</tr>
<tr>
<td>hrv-54</td>
<td>hrv-67</td>
<td>hrv-60</td>
<td>VP2, VP3, VP1, P2A, P2B, P2C, P3A</td>
<td>1.561 × 10⁻²⁸</td>
</tr>
<tr>
<td>hrv-53</td>
<td>hrv-80</td>
<td>hrv-46</td>
<td>5’UTR, VP4, VP2, VP3, VP1</td>
<td>1.291 × 10⁻⁶²</td>
</tr>
<tr>
<td>hrv-53</td>
<td>HRV41</td>
<td>hrv-73</td>
<td>5’UTR, VP4, VP2, VP3, VP1</td>
<td>9.419 × 10⁻¹²</td>
</tr>
<tr>
<td>hrv-53</td>
<td>hrv-59</td>
<td>hrv-39</td>
<td>5’UTR, VP4, VP2, VP3, VP1</td>
<td>2.963 × 10⁻¹²</td>
</tr>
<tr>
<td>hrv-53</td>
<td>hrv-20</td>
<td>hrv-28</td>
<td>VP4, VP2, VP3, VP1, P2A, P2B, P2C</td>
<td>6.839 × 10⁻¹²</td>
</tr>
<tr>
<td>hrv-100</td>
<td>hrv-10</td>
<td>hrv-56</td>
<td>P2A, P2B, P2C, P3A, P3B, P3C, P3D</td>
<td>4.230 × 10⁻¹⁷</td>
</tr>
<tr>
<td>hrv-29</td>
<td>hrv-54-405</td>
<td>hrv-31</td>
<td>P3C, P3D</td>
<td>7.934 × 10⁻¹³</td>
</tr>
<tr>
<td>hrv-29</td>
<td>hrv-54-405</td>
<td>hrv-47</td>
<td>P3C, P3D</td>
<td>1.232 × 10⁻¹⁰</td>
</tr>
<tr>
<td>hrv-42</td>
<td>hrv-97</td>
<td>hrv-4</td>
<td>5’UTR, VP4</td>
<td>8.233 × 10⁻³⁵</td>
</tr>
<tr>
<td>hrv-84</td>
<td>hrv-37</td>
<td>hrv-27</td>
<td>5’UTR, VP4</td>
<td>1.500 × 10⁻¹⁰</td>
</tr>
<tr>
<td>hrv-84</td>
<td>hrv-37</td>
<td>hrv-93</td>
<td>5’UTR, VP4</td>
<td>1.500 × 10⁻¹⁰</td>
</tr>
<tr>
<td>hrv-84</td>
<td>hrv-37</td>
<td>hrv-97</td>
<td>5’UTR, VP4</td>
<td>1.500 × 10⁻¹⁰</td>
</tr>
</tbody>
</table>

Fig. 3. Recombination of HRVs creates additional serotypes. (A) Representative results showing that hrv-46 arose from a recombination of hrv-53 (major parent) and hrv-80 (minor parent). Shown are normalized pairwise identities between each parent and the daughter hrv (purple and green) and the two parents (yellow). As indicated, hrv-46 nucleotides 32 to 3222 are from hrv-80, and nucleotides 3223 to 7200 are from hrv-53. (B) Recombination with an ancestor of hrv-54-f05 has resulted in seven serotype progeny. Each parental hrv is shown as a solid color. The contribution of each parent is proportional to the area of its color in the offspring. See Table 1 and table S4 for nucleotide boundaries and results from other recombination events.
Photodegradable Hydrogels for Dynamic Tuning of Physical and Chemical Properties

April M. Kloxin, Andrea M. Kasko, Chelsea N. Salinas, Kristi S. Anseth

We report a strategy to create photodegradable poly(ethylene glycol)-based hydrogels through rapid polymerization of cytocompatible macromers for remote manipulation of gel properties in situ. Postgelation control of the gel properties was demonstrated to introduce temporal changes, creation of arbitrarily shaped features, and on-demand pendant functionality release. Channels photodegraded within a hydrogel containing encapsulated cells allow cell migration. Temporal variation of the biochemical gel properties was demonstrated to introduce temporal changes, creation of arbitrarily shaped features, and on-demand pendant functionality release.

Hydrogels are hydrophilic polymers swollen by water that are insoluble owing to physical or chemical cross-links. These water-swollen gels are used extensively as biomaterials for complex device fabrication (1), cell culture for tissue regeneration (2), and targeted drug release (3). Often, sophisticated control of the gel structure in space and time is required to elucidate the dynamic relationship between biomaterial properties and their influence on biological function (4, 5). For example, progenitor cells are often expanded and differentiated in hydrogel microenvironments, and researchers have demonstrated how the initial gel properties, including mechanics (6, 7) and chemical functionality (8), influence cellular fate. In regenerative medicine, the structure and composition of gels are also regulated temporally, through hydrolytic (9) and enzymatic (10–12) degradation mechanisms, to promote cell secretory properties and encourage the development of tissue-like structures in vitro and in vivo. A major challenge is determining which biochemical and biophysical features must be presented in a gel culture environment.

Hydrogel structure and functionality have evolved from the direct encapsulation of cells in simple homogeneous materials to those with highly regulated structures spanning multiple size scales (e.g., through self-assembly (13) or microengineering (14)). These hydrogel structures are further modified locally by cells with the synthetic incorporation of bioreponsive functionalities (15) or externally by advanced patterning to create spatially varying functionalities. For example, the chemical patterning of a gel by the addition of a second, interpenetrating network or peptide tether has been demonstrated by diffusing chemical moieties into a gel and covalently linking these functionalities to the network by photocoupling (16) or reaction with a photolytically uncaged reactive group (17). Although these are important advances, such processes do not allow modulation of the gel chemistry in real time or photooxidation of the gel structure. Few synthetic materials provide a cellular microenvironment in which physical or chemical cues are initially present and subsequently regulated on demand. We have synthesized monomers capable of polymerizing in the presence of cells to produce photolytically degradable hydrogels whose physical or chemical properties are tunable temporally and spatially with light. The desired gel property for altering cell function or fabricating a

References and Notes
13. Materials and methods are available as supporting material on Science Online.
19. On the basis of these data, the International Committee on Taxonomy of Viruses Picornavirus Study Group recently proposed to recognize them as a new species (29).
20. hrv-D8 and hrv-95 are designated as two different serotypes by ATCC but differ by 67 nucleotides, perhaps indicating a misidentification in the repository.
29. N. Knowles, personal communication.
30. Funded by the University of Maryland School of Medicine internal funds and by NIH grant U19-A070503. We thank A. Wolf and J.-Y. Sgro for implementation of the tree topology tests. The accession numbers for the hrv strains are FJ445111 to FJ445190 (see table S1).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1165557/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S5
References
5 September 2008; accepted 4 February 2009
Published online 12 February 2009;
10.1126/science.1165557
Include this information when citing this paper.

www.sciencemag.org  SCIENCE VOL 324 3 APRIL 2009 59