Strategy for Systematic Assembly of Large RNA and DNA Genomes: Transmissible Gastroenteritis Virus Model

BOYD YOUNT,1 KRISTOPHER M. CURTIS,2 AND RALPH S. BARIC1,2,*

Department of Epidemiology, Program of Infectious Diseases, School of Public Health,1 and Department of Microbiology and Immunology, School of Medicine,2 University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received 18 May 2000/Accepted 15 August 2000

A systematic method was developed to assemble functional full-length genomes of large RNA and DNA viruses. Coronaviruses contain the largest single-stranded positive-polarity RNA genome in nature. The ∼30-kb genome, coupled with regions of genomic instability, has hindered the development of a full-length infectious cDNA construct. We have assembled a full-length infectious construct of transmissible gastroenteritis virus (TGEV), an important pathogen in swine. Using a novel approach, six adjoining cDNA subclones that span the entire TGEV genome were isolated. Each clone was engineered with unique flanking interconnecting junctions which determine a precise systematic assembly with only the adjacent cDNA subclones, resulting in an intact TGEV cDNA construct of ∼28.5 kb in length. Transcripts derived from the full-length TGEV construct were infectious, and progeny virions were serially passaged in permissive host cells. Viral antigen production and subgenomic mRNA synthesis were evident during infection and throughout passage. Plaque-purified virus derived from the infectious construct replicated efficiently and displayed similar plaque morphology in permissive host cells. Host range phenotypes of the molecularly cloned and wild-type viruses were similar in cells of swine and feline origin. The recombinant viruses were sequenced across the unique interconnecting junctions, conclusively demonstrating the marker mutations and restriction sites that were engineered into the component clones. Full-length infectious constructs of TGEV will permit the precise genetic modification of the coronavirus genome. The method that we have designed to generate an infectious cDNA construct of TGEV could theoretically be used to precisely reconstruct microbial or eukaryotic genomes approaching several million base pairs in length.

Molecular genetic analysis of the structure and function of RNA virus genomes has been profoundly advanced by the availability of full-length cDNA clones, the source of infectious RNA transcripts that replicate efficiently when introduced into permissive cell lines (2, 9). Recombinant DNA technology has allowed the isolation of infectious cDNA clones from a number of positive-stranded RNA viruses, including picornaviruses, caliciviruses, alphaviruses, flaviviruses, and arteriviruses, whose RNA genomes range in size from ∼7 to 15 kb in length (1, 13, 32, 34, 35, 47, 48, 54). The availability of these clones has enhanced our understanding of the molecular mechanisms of viral replication and pathogenesis and resulted in new approaches for heterologous gene expression and vaccine development.

The order Nidovirales includes mammalian positive-polarity single-stranded RNA viruses in the arterivirus and coronavirus families (10, 16). The Coronaviridae family includes the Coronavirus and Torovirus genera (10, 46). Despite significant size differences (∼13 to 32 kb), the polystrictonic genome organization and regulation of gene expression from a nested set of subgenomic mRNAs are similar for all members of the order (16, 46). The family Coronaviridae contains the largest RNA viral genomes in nature (26, 44). Transmissible gastroenteritis virus (TGEV), a group I coronavirus, contains a ∼28.5-kb genomic RNA that is packaged into a helical nucleocapsid structure and surrounded by an envelope that contains three virus-specific glycoprotein spikes, including the S glycoprotein, membrane glycoprotein (M), and a small envelope glycoprotein (E) (17, 18, 33, 36). The TGEV genome is polycistronic and encodes eight large open reading frames (ORFs), which are expressed from full-length or subgenome-length mRNAs during infection (17, 42, 43). The 5’-most ∼20 kb encode the RNA replicase genes, which are encoded in two large ORFs, designated 1a and 1b, the latter of which is expressed by ribosomal frameshifting (3, 17). ORF1a encodes at least two viral proteases and several other nonstructural proteins, while ORF1b contains polymerase, helicase, and metal-binding motifs typical of an RNA polymerase (3, 17, 19). In the 3’-most ∼9 kb of the TGEV genome, each of the downstream ORFs is preceded by a highly conserved intergenic sequence element, which directs the synthesis of each of the six or seven subgenomic RNAs (11, 17, 18, 52). These subgenomic mRNAs are arranged in a nested set structure from the 3’ end of the genome, and each contains a leader RNA sequence derived from the 5’ end of the genome (26, 29, 42, 43). Subgenomic mRNAs are generated by a discontinuous transcription mechanism, the details of which are somewhat controversial (4, 40, 42, 43). In addition to the viral mRNAs, full-length and subgenome-length negative-strand RNAs are implicated in mRNA synthesis (4, 26, 40, 42, 43). Another unique feature of coronavirus replication is the high RNA recombination frequencies associated with infection (6, 25, 26).

The large size of the coronavirus genome, coupled with the inability to clone portions of the polymerase gene in microbial vectors, has hampered the ability to perform precise manipulations and reverse genetics in members of the Coronaviridae (17, 18, 26, 44). Recently these problems were overcome when a full-length cDNA of TGEV was stably cloned in bacterial artificial chromosome (BAC) vectors (3). In this report, we...
describe a simple and rapid approach for systematically assembling a full-length, infectious cDNA construct of TGEV using a series of smaller subclones and a novel strategy which theoretically may allow the assembly of large microbial or eukaryotic genomes approaching several million base pairs in length.

MATERIALS AND METHODS

Viruses and cells. The Purdue strain (ATCC VR-763) of TGEV was obtained from the American Type Culture Collection (ATCC) and passaged twice by swine cell lines. MC21 was passaged in pig kidney cell line. ST cells were obtained from the ATCC (ATCC 1746-CRL) and maintained in minimal essential medium (MEM) containing 10% fetal bovine serum at 37°C. Wild-type TGEV or viruses (icTGEV) derived from the full-length construct were plaque purified twice, and stocks were grown in ST cells as described (42, 43). To measure the growth rate of different viruses, cultures of ST cells seeded at 10^6 cells/ml were infected with 10^5 plaque forming units (pfu) for various molecularly cloned fragments. Cytidine was incorporated to reduce unwanted amino acid changes associated with reverse transcription or naturally occurring quasispecies variation.

Assembly of a full-length TGEV infectious construct. Each of the plasmids was grown to high concentration, isolated, and digested or double-digested with BglI, BstXI, or NdeI according to the manufacturer’s directions (Tel-TEST B., Inc.). To isolate the TGEV subclones, reverse transcription was performed using Superscript II and oligodeoxynucleotide primers according to the manufacturer’s recommendations (Boehringer Mannheim Biochemicals). First-strand cDNA synthesis was at 55°C for 1 h, the cDNA was denatured for 5 min, and annealed by adding 43 U of reverse transcriptase (RT-PCR) using primer pairs directed against the Purdue strain of TGEV or a Taiwanese isolate (11, 17, 26). To provide N protein in vitro transcription reactions. The full-length TGEV construct is designated TGEV 1000. RT reactions were performed using Superscript II reverse transcriptase for 1 h at 50°C as described by the manufacturer (Gibco-BRL) prior to PCR amplification with the reverse primer that flanked a particular interconnecting junction. To amplify across the B1-B2 junction, forward (5'-GCTCAGTGTTTAGCTCAAACAGG-3') and reverse (5'-TCAAGACGACATTTACGCTAG-3') primers were located at nucleotides 9738 to 9759 and 10248 to 10270, respectively, and resulted in a 532-bp amplicon (17). In virus derived from the infectious construct, BsoXI digestion should result in 321- and 211-bp fragments. To amplify across the B2-C junction, forward (5'-TGTCAGGCAGACATTTACGCTAG-3') and reverse (5'-CTCTTGACGACTATTACGCTAG-3') primers were located at nucleotides 11231 to 11251 and 11634 to 11655, respectively, and resulted in an amplicon of 424 bp (17). In molecularly cloned virus, BglI digestion should result in products of 300 and 124 bp. To amplify across the C-D1 junction, forward (GAATGTCACATCAGAGCTC) and reverse (AGCGGTTGTATGTTGTTGCTG) primers were located at nucleotides 16380 to 16400 and 16936 to 16957, respectively (17). If a BsoXI site is present in this 577-bp amplicon, digestion should result in products of 370 and 207 bp. To amplify across the D1-F junction, forward (CTTTGGTTGCTGAACTGAC) and reverse (CTCTCCGGTCTGTTGTTGCTG) primers were located at nucleotides 23034 to 23234 and 23852 to 23873 in the 5 gene, respectively (3, 33). Following BsoXI digestion of this 569-bp amplicon, 586- and 183-bp fragments should be visualized in icTGEV virus. To further confirm the in vitro transcription reactions. Following 28 cycles of amplification with Taq polymerase (Expand Long Kit; Roche Biochemical), the PCR products were separated and visualized by agarose gel electrophoresis. PCR amplicons were then cloned directly into pGEMt cloning vector using BglI digestion and ligated into NEB4. The full-length construct used in this study is designated TGEV 1000.
FIG. 1. Strategy for directionally assembling a TGEV infectious construct. (A) The TGEV genome is a linear positive-polarity RNA of about 28,500 nucleotides. Using RT-PCR and unique oligonucleotide primer mutagenesis, five clones spanning the entire TGEV genome were isolated using standard recombinant DNA techniques. Unique BglI sites were inserted at the junctions between each clone, a unique T7 start site was inserted at the 5' end of clone A, and a 25-nucleotide T
tail and downstream NotI site were inserted at the 3' end of clone F. The approximate location of each site is shown. (B) Cloning the TGEV B amplicon. Because of chromosomal instability in E. coli, it was noted that two B clones (TGEV B1 and B3) contained large insertions at nucleotide 9973 in the TGEV genome (17). Other TGEV B clones had deletions across these sequences. Assuming that these insertions and deletions were “detoxifying” poison TGEV sequences in E. coli, we bisected the B fragment by inserting a BstXI site at position 9949 and cloning two separate clones designated TGEV B1-1,2 and TGEV B2-1,2. Quasispecies variation in the sequence of each independent plasmid clones is shown, with conserved changes denoted with asterisks. These conserved changes differed from the published sequence reported by Eleouet et al. (17) but were identical to the sequence reported by Almazan et al. (3). To reconstruct a wild-type B1 fragment, an SfiI-PflI fragment from B1-1 was isolated and inserted into the TGEV B1-2 backbone to produce a consensus B1 clone.
TABLE 2. Primer pairs for assembly of the TGEV infectious construct. Primers were designed to anneal upstream of conserved sequence motifs and downstream of conserved sequence motifs. The TGEV 3' end of the infectious cDNA construct was generated by RT-PCR using TGEV A, B, C, DE, and F primers (designated TGEV A, B, C, DE, and F). Each cDNA subclone was ligated sequentially into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Orientation Oligonucleotide</th>
<th>Primer in Frame of the TGEV infectious construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGEV 3' end, insert Bgl site</td>
<td>+</td>
<td>2457–2472</td>
<td>3' end, Polya (null)</td>
<td>TGEV 3' end, insert Bgl site</td>
</tr>
<tr>
<td>TGEV 3' end, insert Bgl site plus +</td>
<td>+</td>
<td>2457–2472</td>
<td>3' end, Polya (null)</td>
<td>TGEV 3' end, insert Bgl site plus +</td>
</tr>
<tr>
<td>TGEV 3' end, insert Bgl site minus</td>
<td>+</td>
<td>2457–2472</td>
<td>3' end, Polya (null)</td>
<td>TGEV 3' end, insert Bgl site minus</td>
</tr>
<tr>
<td>Remove T7 stop RT-PCR plus +</td>
<td>+</td>
<td>1029–1049</td>
<td>5' end, insert Bgl site plus +</td>
<td>Remove T7 stop RT-PCR plus +</td>
</tr>
<tr>
<td>Remove T7 stop RT-PCR plus</td>
<td>+</td>
<td>1029–1049</td>
<td>5' end, insert Bgl site plus</td>
<td>Remove T7 stop RT-PCR plus</td>
</tr>
</tbody>
</table>

Theoretical restriction enzymes.

Conventional restriction enzymes, such as Bgl II and EcoRI, leave sticky ends that assemble with the presence of duplications, inversions, and repetitive sequences. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.

Assembly of a full-length TGEV construct. Initially, we isolated five cDNA subclones spanning the entire TGEV genome (Table 2). Each cDNA subclone was ligated precisely into an intact full-length TGEV cDNA construct (Fig. 1A). To RT-PCR clone the 6.2-kb DNA fragment located at the 5' end of the genomic RNA (Fig. 1), we assembled a sequential series of smaller DNA subclones, each of which were cloned into the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites and will only anneal with the appropriate 3-nucleotide overhang-containing TGEV A, B, C, DE, and F primer sites. To address these questions, we determined the frequency of identical-end duplicity, however, can be circumvented, these data suggest that the genomes of many microbial and viral pathogens, including those of microbial and viral pathogens, are devoid of identical-end duplicity.

Theoretical framework.

Theoretical framework.
the forward primer included a T7 start site and the 5'9'-most TGEV leader RNA sequences, while the reverse primer was located at nucleotide 6180, just downstream from a naturally occurring BglI site (GCCTGGTT) in the TGEV genome (3, 17) (Table 1). The 5.2-kb B fragment was amplified using a forward primer upstream of the BglI site at position 6159 and a reverse primer which introduced a unique BglI site (GCCGATGGTCCGG) at position 11355 (Fig. 1). The 5.2-kb C fragment was amplified using a forward primer which introduced the same BglI site at nucleotide 11355 and a reverse primer which introduced another unique BglI site (GCCCTCAGTCGGG) at position 16595. While our original cloning strategy called for separate D and E fragments, it became evident that a single 6.9-kb fragment was stable in microbial vectors. Therefore, a single DE-1 fragment was amplified using a forward primer that introduced the same BglI site at position 16595 and a reverse primer which introduced a new BglI site (GCCGATGGTCCGG) in the S glycoprotein gene at nucleotide 23487. The F fragment was cloned with a forward primer that introduced the same BglI site at position 23487 and a reverse primer that contained the 39'-most nucleotides of the TGEV genome, including an additional 25 T's prior to terminating at a NcoI site. A list of the primers used to mutagenize the TGEV genome and to isolate each of the TGEV subclones

### TABLE 2. Restriction enzymes that cleave at specific sites and leave variable sticky ends

<table>
<thead>
<tr>
<th>Restriction enzyme</th>
<th>Restriction sites</th>
<th>No. of sticky ends</th>
<th>Cutting frequency</th>
<th>Theoretical end redundancy</th>
<th>Actual no. of restriction sites((% \text{ with nonunique ends}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BglI</td>
<td>GCCNNNN</td>
<td>GCNG 2NGGC 34,096</td>
<td>4,096 ± 23,266</td>
<td>262,144</td>
<td>1 (0) 32 (38) 198 (ND) 3 (0) 8 (0) 19 (26) 57 (79) 1 (0)</td>
</tr>
<tr>
<td>BsrXI</td>
<td>CCANNNNN</td>
<td>NTGG GGTN 4 64,596</td>
<td>4,096</td>
<td>1,048,576</td>
<td>3 (0) 31 (6) 80 (41) 16 (13) 103 (ND) 33 (12) 135 (ND) 3 (0)</td>
</tr>
<tr>
<td>SfiI</td>
<td>GCCNNNNN</td>
<td>NGGCC CCGGN 3 65,536</td>
<td>4,096</td>
<td>4,194,304</td>
<td>2 (0) 68 (62) 0 0 2 (0) 0 0</td>
</tr>
<tr>
<td>SapI</td>
<td>GTCCCTCN</td>
<td>NNN CCGAGAAGNNN 3 16,384</td>
<td>4,096</td>
<td>1,048,576</td>
<td>2 (0) 11 (18) 34 (35) 10 (60) 27 (30) 14 (29) 220 (ND) 4 (0)</td>
</tr>
<tr>
<td>EcoRI</td>
<td>G AATTC</td>
<td>CTTAA G 0 4,096</td>
<td>4,096</td>
<td>4,096</td>
<td>3 15 15 71 74 52 283 7</td>
</tr>
</tbody>
</table>

a Mean distance between BglI sites (in base pairs) in the genomes of these organisms (fragment sizes ranged from 9 to 19,414 bp).

b Frequency of end compatibility with a random sequence.

MV, Marburg virus (19,104-bp genome; NC001688); VZV, varicella-zoster virus (124,884-bp genome; X04370); EBV, Epstein-Barr virus (172,281-bp genome; V01555); FP, fowl pox virus (288,539-bp genome; AF198100); MG, M. genitalium (580,074-bp genome; NC000908); HH6, human herpesvirus 6B strain Z29 (162,114-bp genome; AF157706); CJ, C. jejuni (1,641,481-bp genome; NC002163); TGEV, 28,586-bp genome (3). ND, not done.
is shown in Table 1. These primer pairs did not alter the amino acid coding sequence of the virus. The sequence of each unique interconnecting junction is shown in Fig. 1.

The pTGEV A, C, DE, and F clones were stable in plasmid DNAs in *Escherichia coli*. The B fragment, however, was unstable, and only a few slow-growing isolates were obtained, all of which contained deletions or insertions in the wild-type sequence. During two different cloning attempts, a 200- to 300-nucleotide fragment from the *E. coli* chromosome was inserted at position 9973, which is in a region of instability in the TGEV genome noted by other investigators (Fig. 1B) (3, 17). In addition, some clones contained a ~500-bp deletion across this region. We reasoned that breaks in the TGEV B sequence at or around nucleotide 9973 might ablate fragment instability and allow the cloning of these sequences into *E. coli*. Since the TGEV B fragment does not contain a *B*stXI site, we used primer-mediated mutagenesis (C-A change at position 9944) to bisect the B fragment into TGEV B1 and TGEV B2 amplicons with an adjoining *B*stXI (CCATTTCACTTTGG) site located at position 9949 in the TGEV genome (Fig. 1, Table 1). The 4-nucleotide overhang generated by *B*stXI would also provide additional specificity and sensitivity in systematically assembling the TGEV subclones. After these modifications, pTGEV B1 and B2 plasmid subclones which were stable and grew efficiently in *E. coli* were rapidly identified. The location of each of the subclones used in the assembly of the TGEV full-length construct is shown in relationship to important motifs or *cis*-acting sequences in the viral genome (Fig. 2).

Data from our lab and others suggest that sequences in and around the TGEV poliovirus 3C-like protease (3-Clpro) motif are either bactericidal or unstable in microbial vectors (3, 17). Inserts from three to six independent subclones from each fragment were sequenced, and a consensus TGEV subclone was assembled using standard recombinant DNA techniques. The consensus sequence of our Perdue TGEV full-length construct contained 15 amino acid changes and numerous silent changes compared with the published sequence (Fig. 2) (17). These changes were also noted by Almazan et al. (3). Type I termination sites might also prevent efficient in vitro transcription of infectious full-length TGEV transcripts from the construct. Two types of sites are known to cause pausing and/or termination by bacteriophage T7 RNA polymerase (28, 37). A type I termination site consists of a stable stem-loop structure that terminates transcription in adjacent stretches of T residues, while a type II pause site consists of a specific 7-bp sequence (ATCTGGT) (28, 37). Transcription termination will occur when a stretch of T's is located 6 to 8 nucleotides downstream of this sequence (28, 37). Type I stops had prevented transcription of vesicular stomatitis virus full-length negative-strand RNAs in vitro with T7 polymerase (58). To preempt potential problems in the generation of full-length TGEV transcripts,

![FIG. 3. Assembly of the TGEV full-length construct. (A) The various TGEV plasmid DNAs were digested with *B*glI, *B*stXI, or *Not*I, and the appropriate-sized products were isolated from agarose gels as described in the text. The TGEV A and B1 fragments, TGEV B2 and C3 fragments, or TGEV DE-1 and F fragments were ligated at 16°C overnight in separate reactions. Appropriate-sized products were isolated from agarose gels. (B) A+B1, (C) B2+C, (D) DE-1+F. Following purification from agarose gels, the purified products are shown in panel A as well.](http://jvi.asm.org/)

![FIG. 4. In vitro transcription from full-length TGEV constructs. The A-B1, B2-C, and DE-1-F contigs were ligated in vitro as described in the text. (A and B) DNA positions after 8 and 30 h of electrophoresis, respectively. Lane 1, purified A-B1 product; lane 2, purified B2-C product; lane 3, purified DE-1-F product; lane 4, 1-kb ladder; lane 5, in vitro-ligated products; lane 6, high-molecular-weight markers. (C) The in vitro-ligated products were transcribed in vitro, and the products were digested with DNase for 15 min at room temperature and separated in agarose gels. Lane 1, high-molecular-weight DNA markers; lane 2, 1-kb DNA ladder; lane 3, in vitro-ligated TGEV products; lane 4, DNase-treated in vitro transcripts. Arrow indicates high-molecular-weight transcripts.](http://jvi.asm.org/)
putative type I T7 RNA polymerase termination sites (long runs of six T’s) were identified in the TGEV consensus sequence that starts at nucleotide 3632 in the pTGEV A subclone and 13615 in the pTGEV C subclone (3, 17). Mutations were introduced by primer-mediated overlapping PCR mutagenesis without altering the coding sequence. Putative type II pause sites were also identified at nucleotides 17551, 19957, and 23103 in the TGEV genome (3), but did not contain the prerequisite downstream T-rich stretch necessary for efficient T7 termination.

To assemble a full-length cDNA construct of TGEV, plasmids were digested with BglII, BstXI, or NotI, and the appropriate sized inserts were isolated from agarose gels (Fig. 3A). The TGEV A and B1, B2, and C, and DE-1 and F fragments were ligated overnight in the presence of T7 DNA ligase. Systematically assembled products were isolated from agarose gels (Fig. 3B to D), and the TGEV A-B1, B2-C3, and DE-1–F fragments were religated overnight. The final products were purified by phenol-chloroform-isomyl alcohol and chloroform extraction, precipitated under ethanol, and then separated in agarose gels (Fig. 4A and B). Clearly, an appropriately sized full-length TGEV cDNA of about 29 kb in length (TGEV 1000) was generated as well as some assembly intermediates (Fig. 4C).

Transfection and recovery of infectious virus. Synthesis of full-length TGEV transcripts was difficult but resulted in high-molecular-weight RNA product (Fig. 4C). To enhance transfection efficiencies, we tested several different strategies to maximize infectivity of the putative full-length transcripts in vitro. Under identical conditions of treatment, about 10 to 20% of ST cells are efficiently transfected with Sindbis replicons encoding GFP (1), compared with about 60 to 80% of the BHK cells (data not shown). As coronavirus host range specificity occurs primarily at entry and the genomic RNA is infectious in a variety of permissive and nonpermissive cells (5, 14, 41, 44, 51), we reasoned that BHK cells might be more sensitive primary hosts because of the intrinsically higher transfection efficiency. In addition, several reports have suggested that the coronavirus N nucleocapsid protein may function as part of the transcription complex and may influence the translation efficiency of viral mRNAs (7, 15, 49). Because these data suggested that N might enhance the infectivity of full-length transcripts, four different transfection strategies were tested in BHK cells. We transfected BHK cells with TGEV transcripts alone, TGEV plus TGEV N gene transcripts, or just TGEV N transcripts. In a parallel experiment, TGEV and TGEV N gene transcripts were treated with RNase A prior to transfection. Following electroporation, the BHK cells were seeded with 10^6 ST cells to serve as appropriate permissive hosts for progeny virus amplification.

Three days posttransfection, supernatants were passaged into ST cells, and cytopathic effect was observed within 36 h postinfection only with supernatants derived from the TGEV-
gene-transfected cultures. Supernatants were harvested at 48 h postinfection and passaged twice more at 48-h intervals in fresh ST cell cultures. Cytopathic effect typical of TGEV infection was evident at each passage (data not shown). Fluorescent antibody staining with mouse anti-TGEV serum demonstrated that viral antigen was clearly present in each passage (data not shown). Using RT-PCR with primer pairs located within the leader RNA sequence and at the 3' end of the TGEV genome, leader-containing subgenomic mRNA transcripts (mRNAs 6 and 7) encoding the N and hydrophobic membrane proteins were also evident in passage 1 ST cells (data not shown). Furthermore, virus derived from the TGEV 1000 transcripts was diluted and inoculated into ST cells, where plaques developed after 48 h (Fig. 5). No significant differences in plaque morphology were noted between the molecularly cloned recombinant viruses and wild type, suggesting that the cDNA construct did not contain debilitating mutations. Transcripts of TGEV plus N gene treated with RNase A prior to electroporation did not result in the production of infectious virus.

Plaque-purified stocks prepared from passages 1 through 3 of icTGEV (icTGEV-1, icTGEV-2, and icTGEV-3, respectively) were used in growth curves and compared to the parental TGEV strain in ST cells. Cultures were infected with virus at a MOI of 5 for 1 h, and samples were harvested at selected times over the next 44 h. No significant differences in the replication of wild-type TGEV- or TGEV 1000-derived viruses icTGEV-1, icTGEV-2, and icTGEV-3 were noted in ST cells, and all viruses replicated to titers that approached 10⁸ PFU/ml within 28 h (Fig. 6). These data indicate that viruses derived from the infectious cDNA construct had phenotypes indistinguishable from those of wild-type TGEV in swine cells. TGEV efficiently utilizes the feline and porcine aminopeptidase N receptors for docking and entry and can replicate efficiently in feline CRFK cells (14, 51) (data not shown). To determine the host range phenotype of these viruses, cultures of ST and CRFK cells were infected with the molecularly cloned icTGEV-1 or icTGEV-3 virus at an MOI of 5 and fixed at 12 h postinfection. Viral antigen expression was measured by FA (Fig. 7). Efficient virus docking and entry were evidenced by significant levels of antigen expression in swine and feline cells infected with the molecularly cloned viruses. These data demonstrate that the molecularly cloned viruses had a host range phenotype similar to that of the wild type.

Identification of marker mutations. Infectious virus derived from transfected cultures should contain the four unique interconnecting junction sequences used in the construction of

![Graph showing growth curves of plaque-purified molecularly cloned viruses.](image-url)

**Fig. 6.** Growth curves of plaque-purified molecularly cloned viruses. Plaque-purified wild-type TGEV and recombinant TGEV viruses (icTGEV-1, icTGEV-2, and icTGEV-3) derived from the infectious construct were inoculated into ST cells at an MOI of 5 for 1 h at room temperature. The virus was removed, and the cultures were incubated in complete medium at 37°C. Samples were harvested at the indicated times and assayed by plaque assay in ST cells.

![Images showing host range phenotype of molecularly cloned icTGEV.](image-url)

**Fig. 7.** Host range phenotype of molecularly cloned icTGEV. Cultures of ST or CRFK cells (10⁵) were inoculated with molecularly cloned viruses icTGEV-1 and icTGEV-3 at an MOI of 5 for 1 h at room temperature. The inocula were removed, and the cells were incubated in complete medium at 37°C for 12 h. Medium was removed, and the cells were fixed in a 50% methanol-acetone mix, washed, and stained by FA as described in the text. (A) Mock-infected ST. (B) icTGEV-1-infected ST cells. (C) icTGEV-3-infected ST cells. (D) Mock-infected CRFK. (E) icTGEV-1-infected CRFK cells. (F) icTGEV-3-infected CRFK cells.
the infectious TGEV 1000 construct (Fig. 1 and 2). If these noncoding mutations produce a neutral phenotype on virus replication, they should also be stable during passage. Consequently, wild-type TGEV, icTGEV-1, and icTGEV-3 were inoculated into ST cells, and intracellular RNA was isolated at 12 h postinfection. Using RT-PCR and primer pairs that asymmetrically flank each of the B1-B2, B2-C, C–DE-1, and DE-1–F junctions, we amplified products of 400 to 600 bp (Fig. 8). Results using restriction fragment length polymorphism analysis demonstrated that none of the marker mutations were present in the wild-type TGEV genome (Fig. 8A). However, the icTGEV-1 and icTGEV-3 viruses both contained the unique marker mutation profiles used to create the unique BglI and BstXI restriction sites within the TGEV 1000 construct (Fig. 8B and C). The PCR products were subcloned, and the reverse complement of the sequence is shown, demonstrating that the appropriate mutations were present in the viruses isolated from the infectious construct (Fig. 9). Clearly, TGEV 1000 transcripts were infectious and produced virus which contained the appropriate marker mutations. These data illustrate that infectious constructs of coronaviruses can be systematically and precisely assembled from a series of smaller subclones in vitro.

**DISCUSSION**

The complete ~30-kb nucleotide sequence for a number of coronaviruses has been available for about 10 years (8, 17, 22, 27), yet until recently, a full-length infectious clone has not been assembled because of size constraints and regions of coronavirus genomic instability in bacterial vectors, the requirement for a vector system which allows simple reverse genetic applications, and the inability to synthesize full-length transcripts in vitro. Each of these inherent restrictions must be circumvented to assemble infectious coronavirus constructs and at the same time allow easy reverse genetic applications. In a landmark achievement, a full-length TGEV infectious clone was recently engineered into BAC vectors using standard DNA techniques (3). Following DNA transfection into ST cells, full-length transcripts were initially transcribed from a cytomegalovirus (CMV) promoter and then amplified by virus replication in the cytoplasm of the cell. In this paper, we describe a rapid approach to systematically assembling a full-length infectious TGEV cDNA from a panel of six smaller subclones using in vitro ligation. These methods will provide a powerful complementary approach to systematically assemble new large...
cDNAs from a variety of microbial pathogens into BAC or other vectors that stably maintain large DNA inserts (30). Importantly, RNA or DNA genomes which are too large, circular, or unstable in these cloning vectors can still be assembled using this in vitro ligation technique. As coronaviruses contain the largest RNA genome, these approaches should permit reverse genetic studies for all RNA viruses.

Evidence from several experiments demonstrated that transcripts of the TGEV 1000 genomic construct were infectious. Transcripts treated with RNase were not infectious, indicating that infection was likely initiated from the RNA transcripts synthesized in vitro. Medium from transfected cultures could be used to propagate infection, with corresponding cytopathology and viral antigen expression in fresh cultures of cells. Progeny virions formed plaques in monolayers of permissive cells, and plaque-purified molecularly cloned virus grew efficiently to levels equivalent to those of wild-type virus in permissive host cells. The host range phenotypes of molecularly cloned viruses and wild-type virus were similar in vitro, although additional experiments are needed to determine if these viruses utilize the feline aminopeptidase receptor for docking and entry into feline cells (51). Most importantly, plaque-purified virus contained the expected BglI and BstXI marker mutations, providing definitive evidence that transcripts driven from the TGEV 1000 construct were infectious in vitro. The presence of these neutral mutations did not restrict the ability of iTCGEV to replicate efficiently in ST cells.

It is remarkable that two entirely different approaches can be exploited to engineer infectious constructs of large RNA and DNA viruses. Our assembly strategy for coronavirus infectious constructs is simple and straightforward and does not depend on the availability of an existing viral defective interfering cDNA clone as a foundation for building the infectious construct (3). In contrast to infectious clones of other positive-strand RNA viruses (1, 2, 3, 13, 32, 35, 54), the TGEV 1000 construct must be assembled de novo and does not exist intact in bacterial vectors, circumventing problems in sequence instability. This did not restrict its applicability for reverse genetic applications, but rather allowed genetic manipulation of independent subclones, which will minimize the introduction of spurious mutations elsewhere in the genome during recombinant DNA manipulation. Another advantage of our approach is that different combinations of restriction sites can be used that generate highly variable 5' or 3' overhangs of 1 to 4 nucleotides in length, further increasing the specificity and sensitivity of the assembly cascade (Table 2). Because of insert toxicity in E. coli, infectious clones of yellow fever virus and Japanese encephalitis virus were assembled by in vitro ligation from two subclones but used conventional restriction enzymes like BamHI, ApaI, and AatI (34, 48). Our strategy, however, prevents spurious self-assembly of subclones and will provide a strong complementary approach to engineering large RNA or DNA genomes into BAC vectors or other vectors that stably maintain large DNA inserts (3).

It is interesting that in both TGEV infectious constructs assembled to date, sequences in or around the TGEV 3Cpro motif were unstable in E. coli. Our studies, coupled with the findings by Almazan et al. (3), suggest that the unstable sequences can be disabled by bisecting the sequence between nucleotides 9758 and 9949 in the TGEV genome. This information may permit the isolation of larger TGEV A-B1, B2-C, and DE-F subclones and allow the assembly of infectious cDNAs following a single DNA isolation-ligation step. It is not clear whether similar unstable sequences are located at this position in other group 1 and group 2 coronaviruses.

Synthesizing ~29-kb transcripts in vitro is problematic and the greatest impediment to generating infectious RNA from the assembled TGEV 1000 construct. Using a DNA launch platform and transcription of TGEV RNAs from a CMV promoter, transfection resulted in ~36 infectious units/10 μg of DNA (3). Using an RNA launch platform, similar results were obtained in our laboratory. Compared with Sindbis virus replicons encoding GFP, we synthesized ~100-fold less full-length TGEV transcripts in vitro, probably due to the extreme size of the viral genome (data not shown). Using transcripts driven from the ~28.5-kb TGEV full-length construct alone, viral structural gene expression was not noted in 105 cells. In BHK cultures cotransfected with TGEV and N gene transcripts, ~100 to 500 cells per 105 cells expressed viral structural proteins under identical conditions (data not shown). At 16 h posttransfection, little if any structural protein expression was noted in BHK cells electroporated with N gene transcripts alone or transcripts treated with RNase A. This compares with transfection efficiencies of greater than 60% using the 11- to 12-kb Sindbis virus noncytopathic replicons encoding GFP. Although less dramatic, similar problems were reported with the ~13-kb infectious arterivirus cDNA clone (54). These problems may be circumvented somewhat by constructing BHK cell lines that simultaneously express the swine aminopeptidase N receptor and T7 RNA polymerase, allowing RNA transcription and transcription in vivo, and direct selection of progeny virus amplification in susceptible BHK cell lines (1, 14, 51). Alternatively, CMV promoters can be inserted at the 5' end of the TGEV A clone, allowing DNA launch of infectious RNA (3).

In our studies, we could not generate infectious full-length transcripts until the putative T7 polymerase stop signals were removed from the TGEV genome, cytidine was included in agarose gels to reduce UV damage to DNA fragments, and BHK cells were used as recipient hosts (21). At this time, we have no direct evidence that the T stretches in the TGEV A and C fragments might act as T7 termination sites, as the RNA structure in these regions has not been characterized biochemically. Inclusion of capped N gene transcripts during the transfection process also enhanced the infectivity of the TGEV full-length construct in three separate trials. It is not completely clear whether these results were simply serendipitous or whether N transcripts were simply protecting the full-length transcripts from degradation by competitively interfering with RNAse activity in cells or culture medium. The N protein may also protect the genome-length RNA in a ribonucleoprotein structure in the cell, enhance infectivity directly by stabilizing or functioning as part of an intact replication complex (7, 15, 26), or enhance the expression of viral mRNAs (49). Interestingly, TGEV engineered into BAC vectors did not require the presence of nucleocapsid protein to enhance transcript infectivity, suggesting an ancillary role for N transcripts in our system (3).

Prior to these and earlier studies (3), targeted RNA recombination using defective interfering donor RNAs was the best method for introducing precise alterations into the structural genes of the group II coronavirus mouse hepatitis virus, but this approach has been essentially limited to the ~9 kb of the mouse hepatitis virus genome (24, 25, 53). The availability of TGEV infectious constructs will obviously benefit studies of all aspects of TGEV biology and pathogenesis, including analysis of the coronavirus replicase and the somewhat controversial transcription processes which govern expression of the subgenome-length mRNAs (17, 40, 42, 43). The future development of TGEV vaccines and expression vectors is a particularly intriguing application, as the polycistronic genome organization and synthesis of subgenome-length mRNAs may
allow the simultaneous expression of multiple foreign genes (18). It will also be relatively easy to target TGEV to other species by simple replacements of the S glycoprotein gene (14, 25, 51). In contrast to arterivirus expression vectors, the coronavirus intergenic sequences rarely overlap upstream ORFs, simplifying the design and expression of foreign genes from downstream intergenic promoters (11, 17, 52). Several TGEV downstream ORFs also appear to encode luxury functions that can be deleted from the viral genome without affecting infectivity in vitro (18, 29, 56, 57). Finally, the helical TGEV nucleocapsid structure may minimize packaging constraints and allow the expression of multiple large genes from a single construct (18, 26, 36).

The theoretical limits of our technique may approach several million base pairs of DNA and provide a rapid approach for inserting large cDNAs into BAC vectors (20, 30, 45). The systematic assembly method should be appropriate for constructing full-length infectious constructs of other large RNA viruses, including coronaviruses (27 to 32 kb), toroviruses (24 to 27 kb), and filoviruses like the Ebola and Marburg viruses (19 kb) (10, 26, 31). Viral genomes which are unstable in prokaryotic vectors might also be successfully cloned using these methods (9, 34, 48). Moreover, full-length infectious double-stranded DNA genomes of adenoviruses and herpesviruses promise to be a powerful tool in vaccination, gene transfer, and therapy (30, 45, 50, 55). Historically, full-length infectious constructs of these DNA viruses have been generated by ligation of DNA fragments, by homologous recombination (the more widely used method), or as full-length clones in BAC vectors and likely identified the minimal set of genes required for replication. Our approach, however, may provide a significant insight into gene organization, structure, and function of prokaryotic and eukaryotic chromosomes has provided significant insight into gene organization, structure, and function of prokaryotic and eukaryotic chromosomes. We thank Robert E. Johnston, Nancy Davis, Patrick Harrington, Mary Schaad, Mark Denison, and Lawrence Park for helpful discussions and encouragement during the course of these studies. This work was supported by a research grant from the National Institutes of Health (AI 23946).

ACKNOWLEDGMENTS

We thank Robert E. Johnston, Nancy Davis, Patrick Harrington, Mary Schaad, Mark Denison, and Lawrence Park for helpful discussion and encouragement during the course of these studies. This work was supported by a research grant from the National Institutes of Health (AI 23946).

REFERENCES


