VIROLOGY 190, 92–105 (1992)

Genetic Evolution and Tropism of Transmissible Gastroenteritis Coronaviruses

CARLOS M. SÁNCHEZ,* FÁTIMA GEBAUER,* CARLOS SUÑÉ,* ANA MENDEZ,*
JOAQUÍN DOPAZO,† AND LUIS ENJUANES*,[1]

*Centro Nacional de Biotecnología and Centro de Biología Molecular, CSIC-UAM, Universidad Autónoma de Madrid, Canto Blanco, 28049 Madrid, Spain; and †Departamento de Sanidad Animal, Instituto Nacional de Investigaciones Agrarias, Embajadores, 68, 28012 Madrid, Spain

Received December 27, 1991; accepted May 15, 1992

Transmissible gastroenteritis virus (TGEV) is an enteropathogenic coronavirus isolated for the first time in 1946. Nonenteropathogenic porcine respiratory coronaviruses (PRCVs) have been derived from TGEV. The genetic relationship among six European PRCVs and five coronaviruses of the TGEV antigenic cluster has been determined based on their RNA sequences. The S protein of six PRCVs have an identical deletion of 224 amino acids starting at position 21. The deleted area includes the antigenic sites C and B of TGEV S glycoprotein. Interestingly, two viruses (NEB72 and TOY56) with respiratory tropism have S proteins with a size similar to the enteric viruses. NEB72 and TOY56 viruses have in the S protein 2 and 15 specific amino acid differences with the enteric viruses. Four of the residues changed (aa 219 of NEB72 isolate and aa 92, 94, and 218 of TOY56) are located within the deletion present in the PRCVs and may be involved in the receptor binding site (RBS) conferring enteric tropism to TGEVs. A second RBS used by the virus to infect ST cells might be located in a conserved area between sites A and D of the S glycoprotein, since monoclonal antibodies specific for these sites inhibit the binding of the virus to ST cells. An evolutionary tree relating 13 enteric and respiratory isolates has been proposed. According to this tree, a main virus lineage evolved from a recent progenitor virus which was circulating around 1941. From this, secondary lineages originated PUR46, NEB72, TOY56, MIL65, BRI70, and the PRCVs, in this order. Least squares estimation of the origin of TGEV-related coronaviruses showed a significant constancy in the fixation of mutations with time, that is, the existence of a well-defined molecular clock. A mutation fixation rate of $7 \pm 2 \times 10^{-4}$ nucleotide substitutions per site and per year was calculated for TGEV-related viruses. This rate falls in the range reported for other RNA viruses. Point mutations and probably recombination events have occurred during TGEV evolution. © 1992 Academic Press, Inc.

INTRODUCTION

Transmissible gastroenteritis virus (TGEV) belongs to one of the two major antigenic groups of mammalian coronaviruses (Siddell *et al.*, 1982; Spaan *et al.*, 1988). The virus was first isolated in 1946 (Cox *et al.*, 1990a; Doyle and Hutchings, 1946). It is an enteropathogenic coronavirus which replicates in both villus epithelial cells of the small intestine and in lung cells. In 1984, a nonenteropathogenic virus related to TGEV, the porcine respiratory coronavirus (PRCV) appeared in Europe (Pensaert *et al.*, 1986; Callebaut *et al.*, 1988). This virus replicates to high titers in the respiratory tract and undergoes only limited replication in unidentified submucosal cell types of the small intestine (Cox *et al.*, 1990a,b). A virus similar to the European PRCV has been recently described in North America (Wesley *et al.*, 1990b). In contrast to TGEV, PRCV exhibited no clinical signs of disease (Pensaert *et al.*, 1986; Duret *et al.*, 1988; Wesley *et al.*, 1990b).

The antigenic cross-reaction among isolates of TGEV and PRCV has been clearly documented (Callebaut *et al.*, 1988; Garwes *et al.*, 1988; Sánchez *et al.*, 1990; Rasschaert *et al.*, 1990; Wesley *et al.*, 1990b). Both types of viruses have common antigenic determinants in the three structural proteins: spike (S), membrane (M), and nucleoprotein (N). The absence of two antigenic sites in the S protein of the PRCV isolates has been the base for their differentiation from the enteric viruses (Sánchez *et al.*, 1990). Sequencing of the S gene of a French PRCV isolate (Rasschaert *et al.*, 1990), and of a 200-nucleotide (nt) fragment of the S gene of a North American PRCV isolate (Wesley *et al.*, 1990a) has revealed that both S proteins contain, at comparable locations within the protein, a single deletion of 224 and 227 amino acids, respectively. These isolates also showed deletions which were different in each virus in the genes coding for the nonstructural proteins, mapping downstream to the 3′-end of the S gene (Britton, 1990; Rasschaert *et al.*, 1990; Wesley *et al.*, 1991). PRCV was transmitted by aerosols and has now been detected in most European countries (Enjuanes and Van der Zeijst, 1992). It has been proposed (Enjuanes and Van der Zeijst, 1992) that PRCV behaves as a natural vaccine against TGE, which makes the study of its origin and evolution interesting. The analysis of the genetic relationship among these respiratory isolates and others with respiratory tropism

[1] To whom reprint requests should be addressed.

0042-6822/92 $5.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

would allow us to determine the molecular basis of their tropism and evolution.

In this manuscript we describe the genetic homology among eight respiratory and three enteric isolates of the TGEV antigenic cluster, which identified amino acids potentially involved in receptor binding sites and conserved areas of the S gene. Based on these viral sequences, an evolutionary tree and mechanisms for TGEV evolution have been proposed.

MATERIALS AND METHODS

Cells and viruses

All viruses were grown in swine testicle (ST) cells (McClurkin and Norman, 1966). The characteristics of the viruses are described in Table 1. For simplicity, the viruses are named in the text with three letters indicating their geographical origin or classical name, followed by two numbers indicating the earliest year of isolation as reported in the literature. The antigenic characteristics of most of these viruses have been previously reported (Sánchez *et al.*, 1990). Viruses were purified as described (Correa *et al.*, 1990).

Virus proteins

Protein analysis was performed after dissolution (1 μg/20 μl) in 0.1 *M* sodium acetate, pH 7, 0.5% sodium dodecyl sulfate (SDS), 1 μM phenylmethylsulfonyl fluoride (PMSF), 0.1 μM *N*-α-p-tosyl-L-lysine chloromethyl ketone (TPCK), and 1 μg/ml pepstatin. When indicated, proteins were deglycosylated by incubation overnight at 37° with protein *N*-glycosidase F (0.04 U/ml, Boehringer-Mannheim), and the reaction was stopped by freezing. Protein were subjected to SDS–7.5% polyacrylamide gel electrophoresis (PAGE) after the samples were reduced with 5% 2-mercaptoethanol (Laemmli, 1970). Finally the proteins were detected by silver staining (Ansorge, 1985).

RNA sequencing

RNA was extracted from purified virions as described by Gebauer *et al.* (1991). RNA was sequenced by oligodeoxynucleotide primer extension and dideoxynucleotide chain termination procedure (Sanger *et al.*, 1977) using the protocol described by Fichot and Girard (1990). For RNA sequencing, primers complementary to the S gene (Gebauer *et al.*, 1991) were used. Sequence data were assembled using the computer programs of the Genetics Computer Group (University of Wisconsin).

Evolutionary tree

Sequence information has been analyzed following standard phylogenetic methods. The distance between each pair of nucleotide sequences was estimated using the formula $d = -(\frac{3}{4})\ln(1 - 4p/3)L$ (Jukes and Cantor, 1969), where p is the proportion of changed nucleotides displayed by the compared sequences, and L is the length of the sequences after alignment. The two gaps introduced to align the sequences were excluded from the calculations. The neighbor-joining method (Saitou and Nei, 1987; Sourdis and Nei, 1988), as implemented in the program TREEDIST (available from J.D. upon request), was used to obtain a phylogenetic tree from the pairwise distance matrix. A parallel phylogenetic analysis was carried out using the least squares method (Fitch and Margoliash, 1967), utilizing the program FITCH from the PHYLIP package, version 3.3 (Felsenstein, 1990). The reliability of the tree, i.e., the confidence levels for branching order, was determined by the bootstrap method (Efron, 1982; Felsenstein, 1985). A high number of bootstrap replicates of the original set of sequences was obtained. For each replicate a phylogenetic tree was obtained as described above. Hence, a consensus topology for the tree, as well as confidence intervals for each branching point (Felsenstein, 1985) were obtained by applying the program CONSENSE, also from the PHYLIP package. Automatized derivation of bootstrap replicates, distance matrices, and neighbor-joining tree estimations were provided by the TREEDIST program.

The origin of the phylogenetic tree was estimated by a lineal least squares procedure (Sokal and Rohlf, 1981). We assumed a constant average rate of fixation of mutations. This procedure determines the origin, finding the point in the tree that minimizes the sum of the squares of a lineal least squares fit, and relates the distances between each isolate and this point to isolation dates. The slope of the line provides an estimate of the rate of fixation of mutations. The interception of the line with the horizontal axis (time) gives an estimate of the origin of the TGEV antigenic cluster of viruses. Errors and confidence intervals were calculated for the slope and the intercept with the time axis (Sokal and Rohlf, 1981).

RESULTS

Structural proteins of enteric and respiratory porcine coronaviruses

Both enteric and respiratory TGEVs have been studied. The respiratory viruses could be grouped in two clusters, one lacking antigenic sites B and C (the

TABLE 1

CORONAVIRUSES USED IN THIS PUBLICATION

Designation	Origin (year of isolation)	Dominant tropism	Characteristics	Reference
TGEV				
PUR46-MAD-CC120	Purdue University, Indiana (1946)	Enteric & Respiratory	Enteric virus originally isolated by Bohl. Passaged 120-fold on ST cells. Reference clone used in our laboratory.	Bohl *et al.*, 1972 Sánchez *et al.*, 1990 Gebauer *et al.*, 1991
PUR46-PAR-CC120	idem	idem	Same origin as PUR46-MAD-CC120 Clone used by H. Laude's group.	Bohl *et al.*, 1972 Rasschaert and Laude 1987
PUR46-UTR-CC120	idem	idem	Same origin as PUR46-MAD-CC120 Clone used at Utrecht University	Bohl *et al.*, 1972 Jacobs *et al.*, 1987
MIL65-AME	Ohio (1965 or before)	idem	Virulent. Passed *in vivo* Plaque purified three times on ST cells.	Wesley, 1990
BRI70-FS772	England (1970)	idem	Maintained by passage in primary cultures of thyroid cells	Garwes *et al.*, 1978
NEB72-RT	Nebraska (1972)	Respiratory	Isolated from the lungs of a healthy adult pig. Passaged in the lungs of gnotobiotic pigs. Passaged in vitro in lung cells and on ST cells.	Underdhal *et al.*, 1974 This manuscript
TOY56-CC168	Japan (1956)	Respiratory (sporadically isolated in enteric tissues)	Received at passage 163 in swine kidney cells. Passaged 5 times on ST cells.	Furuuchi *et al.*, 1976 Sánchez *et al.*, 1990
PRCV				
HOL87-V78-CC5	The Netherlands (1987)	Respiratory	Originally isolated on ST cells and passaged 5 times on this cell line	Pensaert *et al.*, 1986 Sánchez *et al.*, 1990
BEL85-83-CC3	Belgium (1985)	idem	idem	idem
BEL87-31-CC5	Belgium (1987)	idem	idem	idem
FRA86-RM4	France (1986)	idem	idem	Duret *et al.*, 1988 Rasschaert *et al.*, 1990
ENG86-I-CC5	England (1986)	idem	Isolate PVC-135308 originally grown on primary pig kidney cells and passaged 5 times on ST cells	Brown and Cartwright, 1986; Garwes *et al.*, 1988; Sánchez *et al.*, 1990.
ENG86-II-CC5	England (1986)	idem	Isolate PVC-137004, isolated and passaged as PVC-135308	idem

PRCVs) and another with these antigenic sites (NEB72 and TOY56) (Sánchez *et al.*, 1990). The molecular weight of the structural proteins of the TGEVs and PRCVs listed in Table 1 were determined, with the exception of those from the isolates BRI70 and FRA86, which have not been analyzed in this study. These molecular weights were estimated by SDS–PAGE analysis. The mobility of the M and N proteins of all viruses was similar (data not shown). In contrast, the TGEV S glycoproteins and the apoproteins, obtained by deglycosylation with protein *N*-glycosidase, had higher molecular weight (200 and 158 kDa, respectively) than the S glycoproteins and apoproteins of the PRCVs (170 and 130 kDa, respectively). The results for the stan-

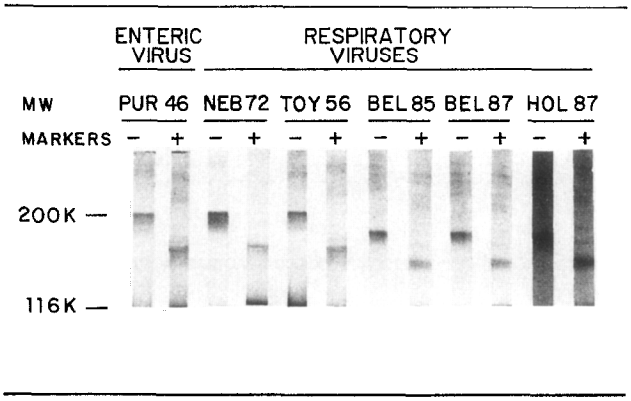


FIG. 1. PAGE analysis of the spike protein of TGEV-related coronaviruses before and after deglycosylation. Purified viruses were dissociated (1 μg/20 μl) in 0.1 *M* sodium acetate, pH 7, with 0.5% SDS and protease inhibitors, and incubated overnight at 37° in the presence (+) or absence (−) of protein *N*-glycosidase F (0.04 U/μl). The proteins were separated by 7.5% PAGE in the presence of 0.1% SDS and 2-mercaptoethanol and detected using silver staining (Ansorge, 1985). Only the gel area corresponding to the S glycoprotein is shown.

dard PUR46 strain, two TGEV strains (NEB72 and TOY56) with respiratory tropism, and three PRCVs (BEL85, BEL87, and HOL87 are shown as representative data (Fig. 1). These results indicate that all the European PRCVs studied have an S protein of similar molecular weight (170 kDa), which is smaller than the S glycoprotein of TGEV. In addition, they demonstrate that other isolates with an almost exclusive respiratory tropism (NEB72 and TOY56) do not have a reduction in molecular weight as were detected in the PRCV isolates.

Sequence analysis of the S-glycoprotein of TGEVs and PRCVs

To determine the relationship among the different European PRCV isolates, the complete S gene sequence of PRCV HOL87, TOY56, and NEB72 respiratory isolates were determined by sequencing the RNA from purified virions (Fig. 2). In addition, the first 1956 nt of the S gene of other four PRCV strains were determined (Fig. 2). The nucleotide or amino acid positions reported in this manuscript refer to the location of equivalent residues in the sequence of MIL65 virus, which has the largest S gene reported for TGEV-related isolates. The 5′-terminal segments sequenced codes for the four antigenic sites previously defined, which are located in the globular part of the peplomer (Gebauer *et al.*, 1991). The sequences were aligned with those of the PRCV FRA86 (Rasschaert *et al.*, 1990) and of prototype enteric viruses (Fig. 2). Two deletions were observed which have been diagrammatically summarized in Fig. 3A. One of them removed 224 amino acids, starting at residue 21 of the unprocessed glycoprotein. The second deletion removed 2 amino acids after residue 374. Taking into account the two deletions and the sequence homology among the S genes of these isolates, three sets of viruses could be distinguished: (i) one including BEL85, FRA86, HOL87, BEL87, ENG86-I, and ENG86-II strains with a 224-aa deletion which was identical both in terms of the number of residues deleted and the location of the deletion; (ii) a second set including PUR46 and NEB72 isolates with a deletion of 2 aa; and (iii) a third set grouping MIL65, BRI70, and TOY 56 strains, which had no deletion. Although the NEB72 and TOY56 isolates have respiratory tropism, they do not contain the 224-amino-acid deletion. These viruses have point mutation differences with the enteric viruses (Fig. 2). The NEB72 isolate has only two amino acid differences when compared to the PUR46 isolate in the S protein, not shown by other enteric isolates. One of them (aa 219) falls within the deletion present in the PRCVs. NEB72 isolate is closely related to PUR46 strain since, in addition, both viruses have the 2-aa deletion (residues 375 and 376) and almost identical sequences in the ORFs 3, 3-1, and 4, corresponding to nonstructural proteins (data not shown). The TOY56 isolate has three amino acid changes (residues 92, 94, and 218) within the deletion present in the PRCVs, in relation with the PUR46 strain, which are specific for the TOY56 isolate. The enteric isolates BRI70-FS and MIL65-AME have also a change in aa 218, from valine to threonine, which is different than the change to isoleucine that occurred in the TOY56 isolate.

The amino acid homology between the S protein of PRCVs and TGEVs was independently studied at the globular and the stem portion of the molecule (data not shown). The same overall degree of homology in the S proteins was found in both the globular and stem areas. The amino acid homology was higher than 98% among both the TGEV and the PRCV isolates. In contrast, the overall S protein homology between TGEVs and PRCVs was around 1% lower. Although this percentage difference is small, the fact that these viruses have the amino acids changed in almost identical location, makes this difference significant. In these comparisons, only the S protein segments for which the sequences of the 13 viruses were available have been considered. A large conserved domain was identified in the globular portion of the S protein of TGEVs and PRCVs, between amino acids 405 and 465, when the number of amino acid changes was plotted versus their position in the sequence (Fig. 3B). Furthermore no amino acid changes were detected in this segment

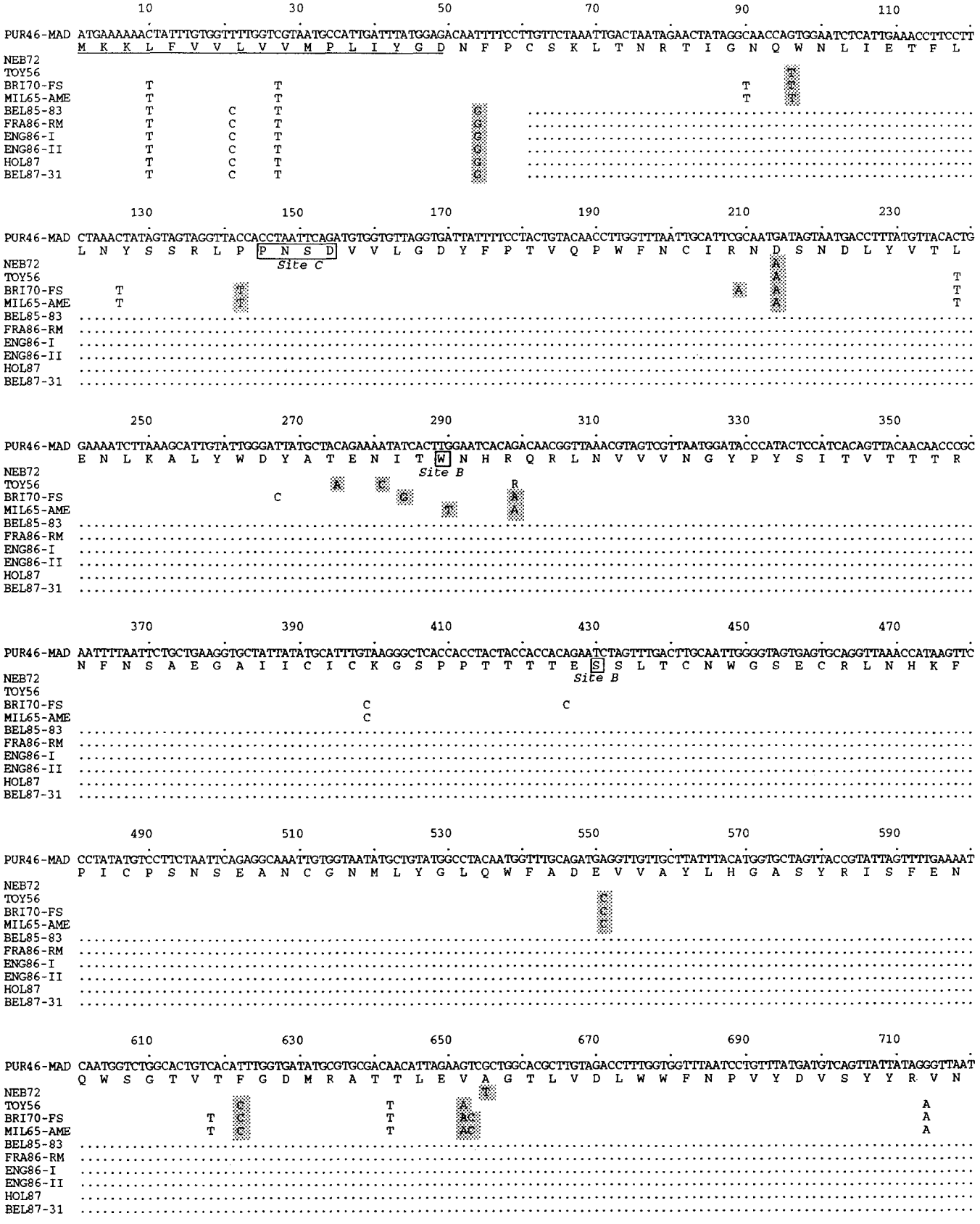
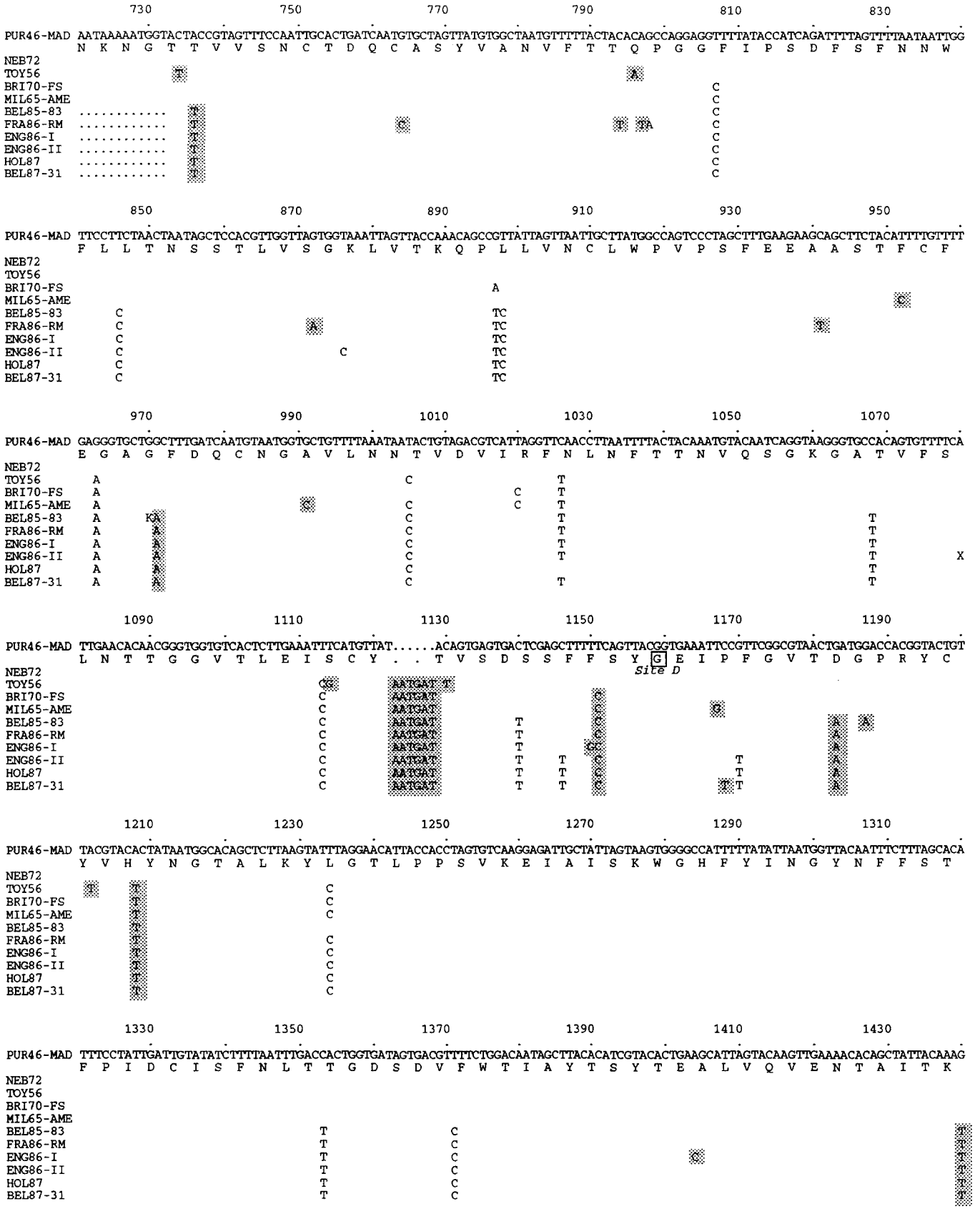


FIG. 2. Sequence alignment of spike (S) protein genes of TGEVs and PRCVs. The nucleotide sequence of the S gene and the deduced aa sequence of the PUR46-MAD virus are shown in the two first lines. In the other lines, the nucleotide changes in the sequences of other viruses have been indicated. Nucleotide changes resulting in amino acid changes have been shadowed. In the alignment deleted residues have been filled out with points. Sequence numbers indicate the positions that the residues would have in the MIL65 virus. For simplicity, the sequences of two clones of the PUR46 isolate (PUR46-PAR and PUR46-UTR) have been omitted in this series of sequences, since they show minor changes and their sequences were previously published. The sequences of the strains FRA86-RM, MIL65-AME, BRI70-FS, PUR46-PAR, and PUR46-UTR



have been previously reported (Britton and Page, 1990; Jacobs *et al.*, 1987; Rasschaert and Laude, 1987; Rasschaert *et al.*, 1990; Wesley, 1990). Sequence indeterminations have been coded as: K for G or T; X for G, A, T, C, or any amino acid; S for C or G; and Y for C or T. Underlined amino acids correspond to the signal peptide. Residues in boxes are involved in the indicated antigenic sites. Asterisks indicate the 3′-end of the segments sequenced. Dashes indicate nonsequenced segments. The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession numbers: PUR46-MAD, M94101; NEB72-RT, M94099; TOY56, M94103; HOL87, M94097; BEL85-83, M94096; BEL87-31, M94098; ENG86-I, M94100; ENG86-II, M94102.

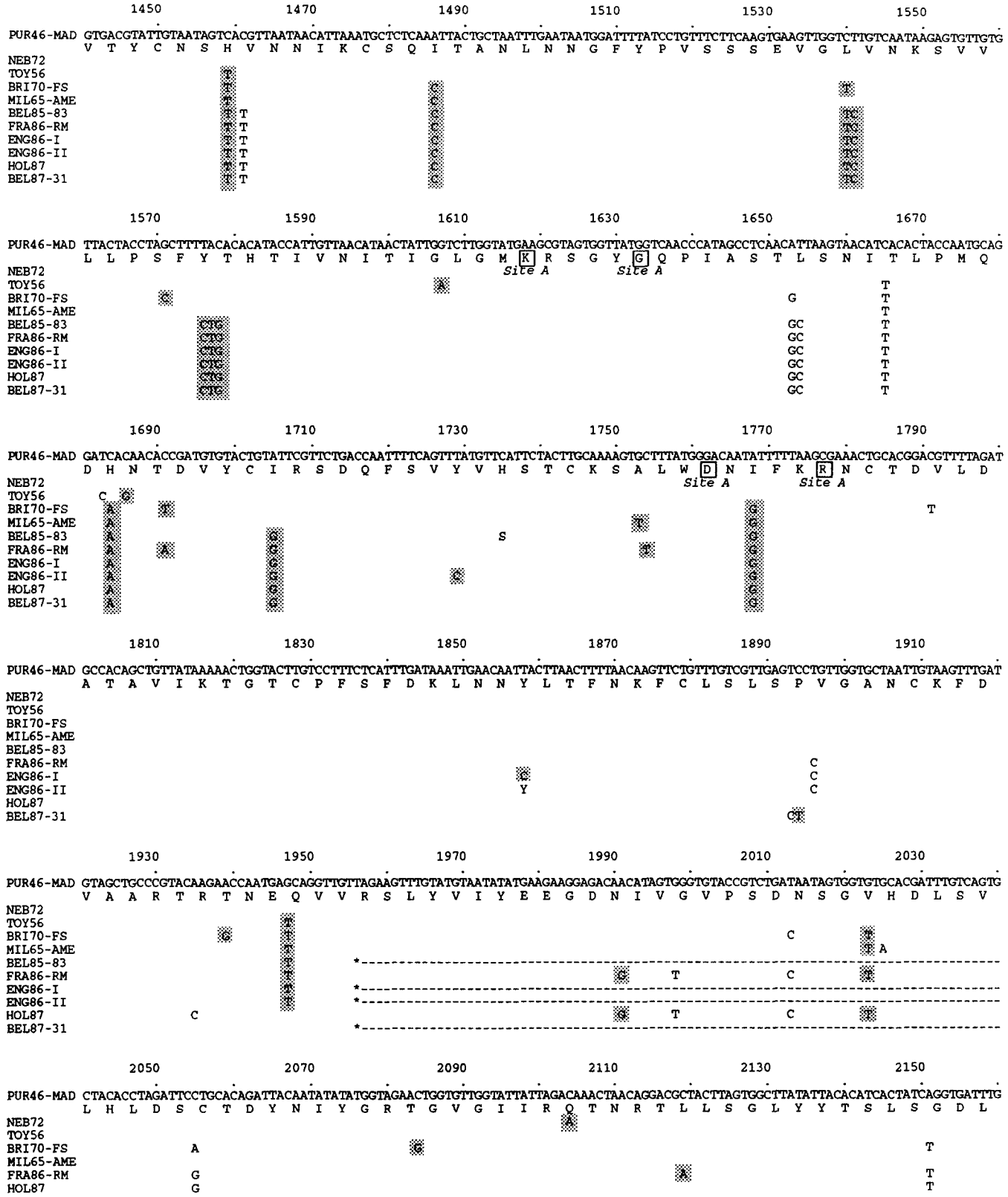


FIG. 2—*Continued*

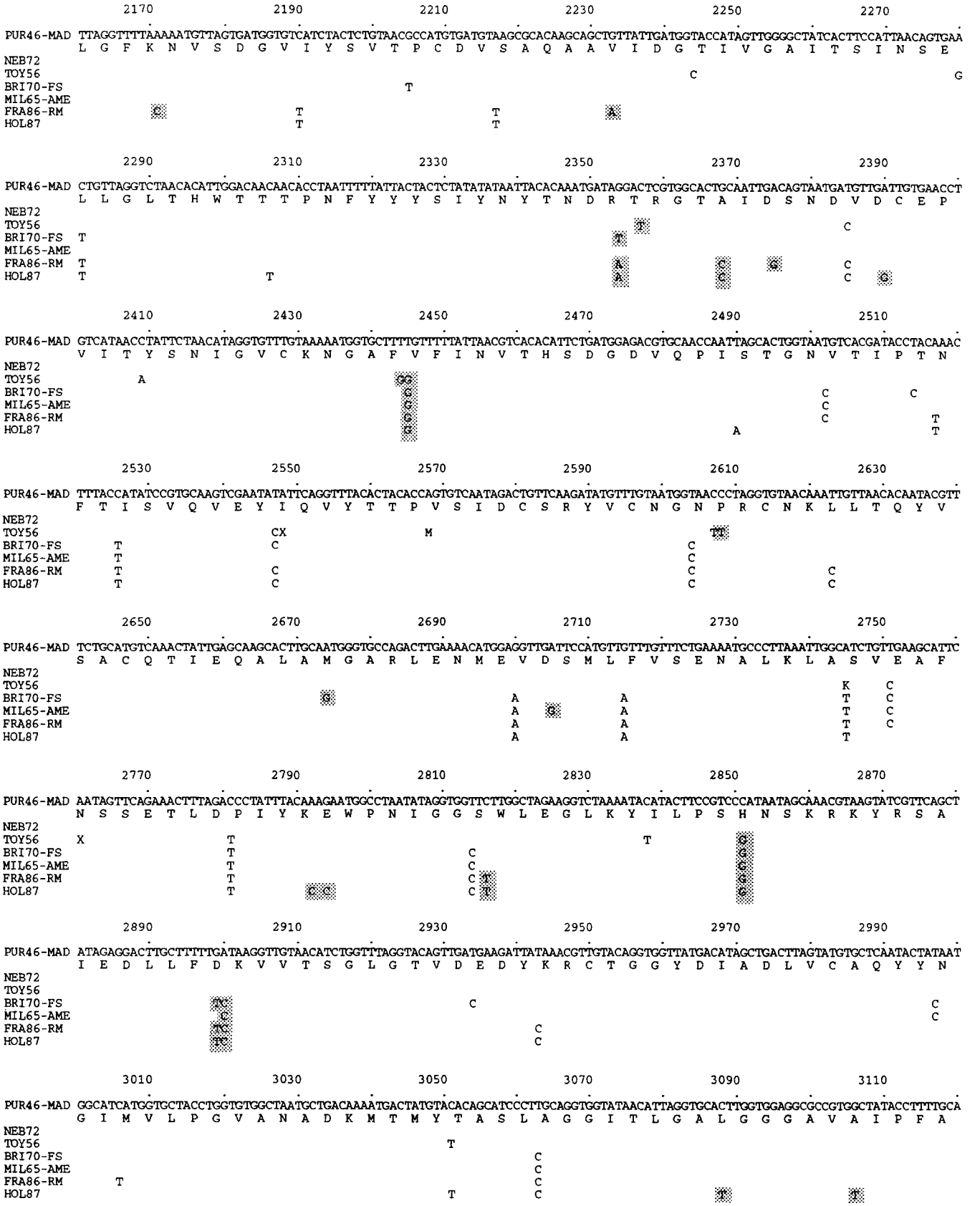


FIG. 2—*Continued*

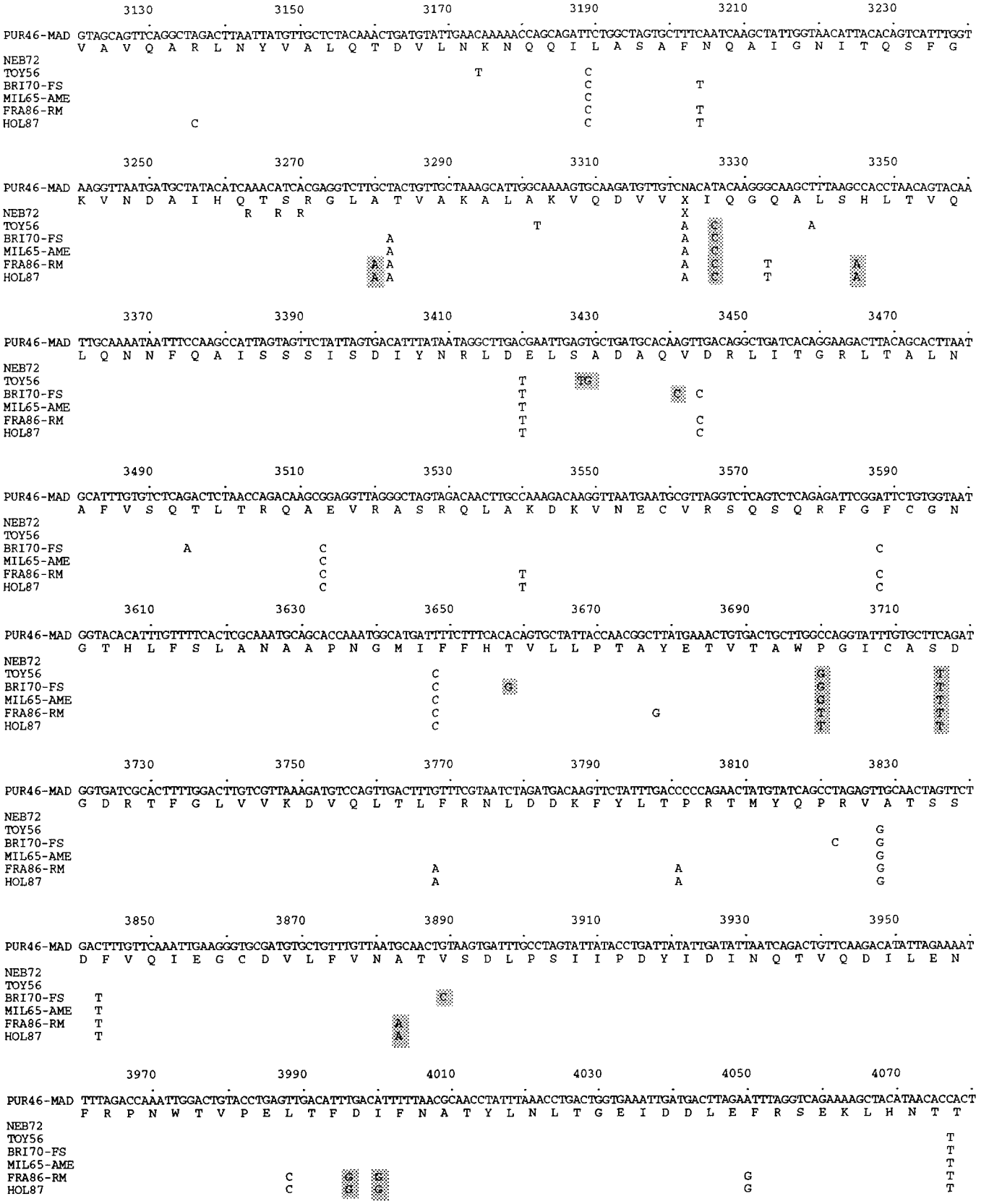


FIG. 2—*Continued*

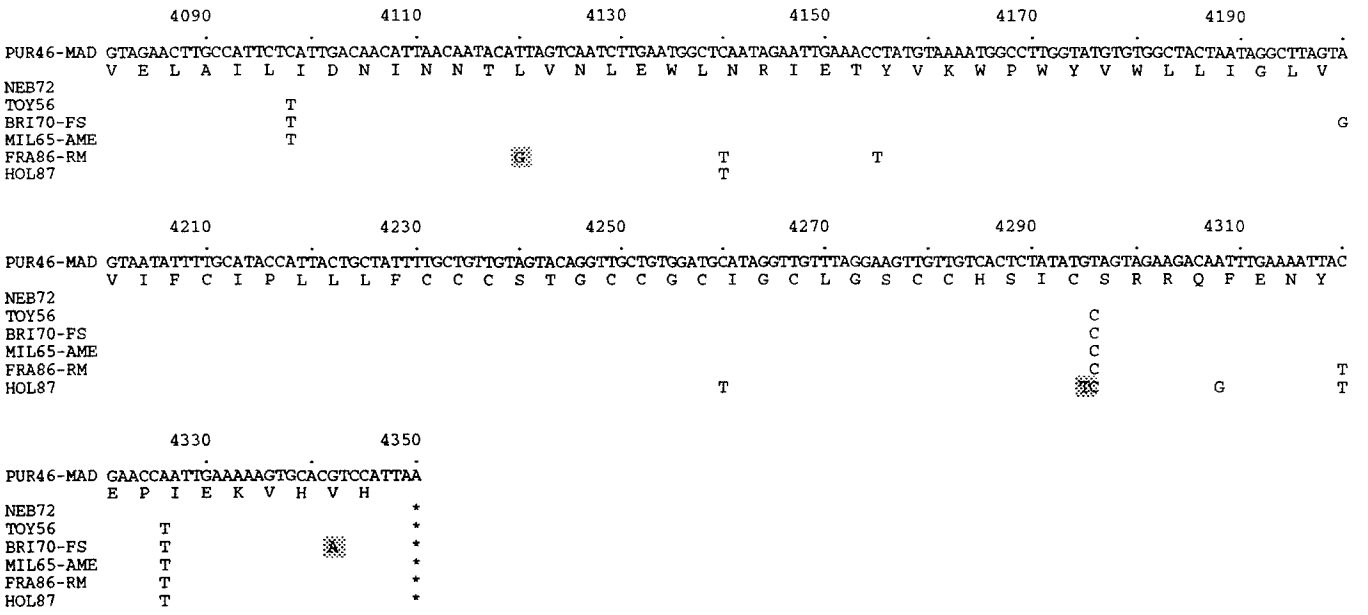


FIG. 2—*Continued*

when the sequences of 13 virus isolates were compared (Fig. 3B).

Evolutionary tree for the S gene of TGEVs and PRCVs

The nucleotide sequence of the S glycoprotein of eight respiratory and five TGEVs (three of which were different clones of the same PUR46 virus strain) were aligned taking into account the two deletions of 6 and 672 nts present in the sequence of the PUR46 and PRCVs, respectively, for maximum fitness (Fig. 2). Phylogenetic analysis of the sequences (first 1956 nt) of the viruses described in Fig. 2, by either the neighbor-joining or the least squares methods of tree-reconstruction procedures, gave two identical trees, with the same branching order, confidence levels, and branch lengths (Fig. 4). This congruence in the results, in addition to the high confidence level along the tree, suggests a significant reliability for the evolutionary history described. The least squares relationship between the number of mutations from origin and the year of isolation was determined (Fig. 5). The extrapolation of this line to zero mutations allowed to predict that these TGEV were originated from a recent common ancestor circulating around 1941. Since then, from a main lineage, the PUR46, TOY56, MIL65, BRI70, and the PRCVs were derived in the indicated order (Fig. 4). Only one isolate (NEB72) accumulated a number of substitutions smaller than the one expected for its year of isolation. In at least three cases (TOY56, MIL65, and BEL85), it can be assured with a significance of 99.9%, that these were lateral lineages derived from one main lineage (see Discussion). The accumulation of mutations with time (Fig. 5) fits a straight line with a high Pearson coefficient correlation ($r^2 = 0.97$). From the slope of this line, the mutations fixation rate can be estimated at 0.95 ± 0.05 substitutions per year.

DISCUSSION

The structural proteins of seven new strains of the TGEV cluster with enteric and respiratory tropism have been analyzed. Also, the complete sequences of the S genes of three respiratory isolates and of the first 1956-nt S gene of other four respiratory viruses of the TGEV antigenic cluster have been determined. These sequences, together with published ones of enteric and respiratory TGEV isolates, have been analyzed to determine the genetic homology between TGEVs and PRCVs. Key point mutations which might be responsible for the loss of enteric tropism in certain isolates have been identified. In addition, a large conserved area in the S protein has been identified, and an evolutionary tree relating all these viruses has been proposed.

TGEVs were described for the first time in 1946 (Doyle and Hutchings, 1946). Respiratory variants of the enteric virus were isolated in 1956 (TOY56 strain) (Furuuchi *et al.*, 1976) and in 1972 (strain NEB72) (Underdahl *et al.*, 1974). Highly contagious respiratory isolates which rapidly extended throughout Europe, the PRCVs, were detected for the first time in 1984 (Pensaert *et al.*, 1986). These viruses are serologically related to TGEV and are missing antigenic sites B and C

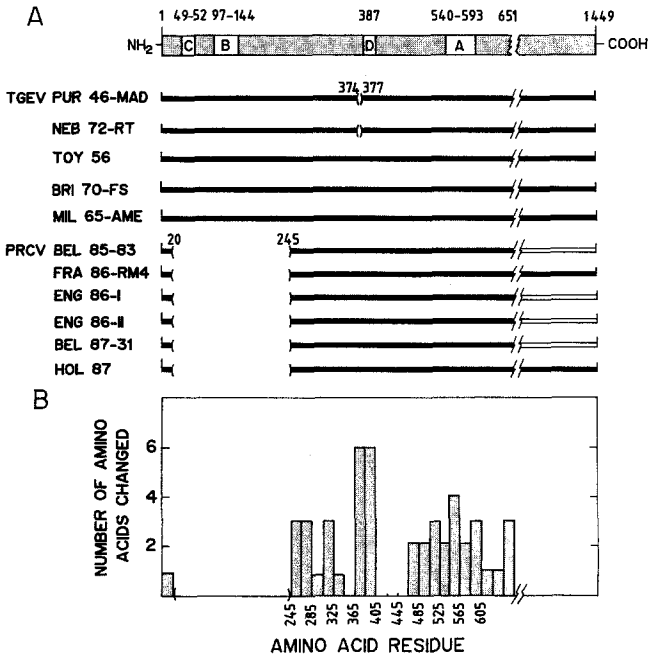


FIG. 3. Summary of the deletions and amino acid changes present in the S glycoprotein of TGEVs and PRCVs. (A) Location of the deletion. Full and empty bars indicate the sequences known and undetermined, respectively. Letters indicate the approximate location of the antigenic sites. The numbers above these letters indicate amino acid residues involved in the formation of these sites. The position of the deletions is indicated by brackets, and the numbers next to the brackets show the amino acids flanking the residues deleted. (B) Number of amino acid changes in sequential fragments of 20 aa each, in relation with the PUR46-MAD virus sequence. Only the segments for which the sequences of the 13 virus strains were available have been included in the comparison. Amino acid residues have been numbered according to their position in the MIL65 virus after the alignment. The origin of the sequences of the different strains has been indicated in the legend of Fig. 2.

(Callebaut *et al.*, 1988; Sánchez *et al.*, 1990). In the five European isolates sequenced by us, the absence of these sites is due to a deletion of 224 amino acids, starting at residue 21. Identical deletion (both in terms of size and location) was described for another European isolate (Rasschaert *et al.*, 1990). In 1989 a virus (IND89) with the antigenic characteristics of the European PRCVs was isolated in the United States (Wesley *et al.*, 1990b). Sequencing of the first 200 nt of the S gene showed a deletion of 227 amino acids starting at residue 23 (Wesley *et al.*, 1991), i.e., the deletion shifted downstream two residues, in relation to the position of the deletion described for the European PRCVs. These data, together with the high sequence homology (Figs. 2), and the phylogenetic tree obtained (Fig. 4) demonstrate that all six European PRCVs, isolated in four countries (Belgium, France, The Netherlands, and United Kingdom) have a recent common ancestor. In contrast, the North American isolate is probably of independent origin, since (i) if it was derived from the European PRCVs the addition of several nucleotides after nt 59 or 60 and a deletion of a few nucleotides at the end of the deletion present in the European PRCVs would have been required (the identity of the nucleotides at the beginning and at the end of the deletion leaves open the precise position of the deletion both in the European and in the North American PRCVs); (ii) differences between the genes coding for the nonstructural proteins of the European and North American isolates have also been reported (Rasschaert *et al.*, 1990; Wesley *et al.*, 1991); and (iii) the 200-nt sequence available for the North American isolate placed this PRCV strain closer to the enteric isolates than to the European PRCVs in our evolutionary tree (results not shown).

The four antigenic sites described in the S glycoprotein of TGEV have been mapped into the NH_2-terminus half of S protein (Delmas *et al.*, 1990; Gebauer *et al.*, 1991). These sites are probably located in the globular part of the S molecule (De Groot *et al.*, 1987; C. Suñé, M. Nermut, J. L. Carrascosa, and L. Enjuanes, unpublished results). In other coronaviruses the S glycoprotein can be split into two subunits, S1 and S2, which probably contain the globular and stem portions of the molecule, respectively (Spaan *et al.*, 1988). The precise residue where the stem part of the S peplomer might start awaits elucidation of its atomic structure. The protein domain that includes antigenic subsites Aa and Ab and site D (Gebauer *et al.*, 1991) showed a slightly higher number of amino acid changes than did other areas of the S protein (Fig. 3). Nevertheless, the overall homology in the globular and stem areas of the S protein is similar in both nucleotide and amino acid levels (results not shown). This result was not anticipated due to the higher antigenicity and presence of epitopes relevant in virus neutralization in the globular area.

The receptor binding site in the S glycoprotein of TGEV that interacts with ST cells probably maps between sites A and D since TGEV binding to ST cells is best inhibited by MAbs specific for these sites (Suñé *et al.*, 1990). Candidate domains for the localization of this receptor binding site could be the highly conserved area identified between amino acids 405 and 465 (Figs. 2 and 3), although other domains around this area can not be ruled out. A second RBS may be used to infect enteric cells by TGEVs with enteric tropism. This RBS might be located within the area of the S protein deleted in the PRCVs. More precisely, it could be located around either aa 92, 94, and 218 or aa 219, changed in the TOY56 or in the NEB72 isolates, respectively. Interestingly, both viruses have an amino

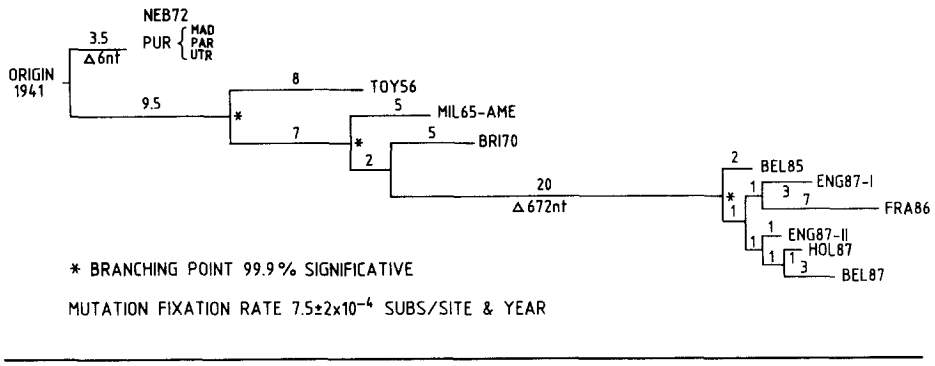


FIG. 4. Evolutionary tree of TGEV related coronaviruses. Neighbor-joining and least squares methods of tree reconstruction procedures were applied to the first 1956 nt of 13 virus isolates (the 11 isolates indicated in Fig. 2, and the clones PUR46-PAR and PUR46-UTR previously reported) (Table 1). Numbers in the diagram indicate residue substitutions between branching points. Δ indicates the introduction of a deletion between branching points. *indicates that all the descendents of this fork have, with a probability of 99.9%, a recent common ancestor.

acid change in contiguous residues (218 and 219), suggesting that these residues may be involved in the RBS. Tissue-specific tropism of coronaviruses is conditioned by the S glycoprotein, and different RBSs in this protein could be recognized by the respiratory or the enteric tissues. Nevertheless, other viral or cellular regulatory mechanisms affecting essential steps of virus replication, other than virus-to-cell binding, could influence viral tropism (Levine, 1984). Genes controlling those regulatory mechanisms could map to areas away from the S gene. Studies based on recombination between TGEVs with enteric and respiratory tropism will help to identify the existence of these genes.

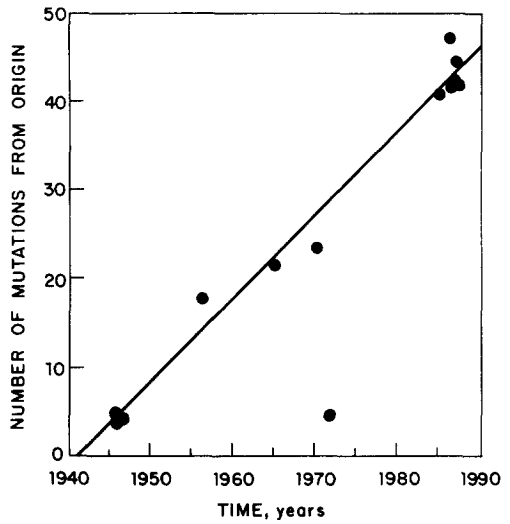


FIG. 5. Relationship between mutation fixation rate and year of isolation. The line relating the number of mutations from origin with the year of isolation was plotted. Line and origin were estimated at the same time by linear least squares fit. The expression for the line was: $d = 0.95\ t - 1893$, $r^2 = 0.97$, where d is the distance to the origin, t is the time in years, and r^2 the Pearson's correlation coefficient (Sokal and Rohlf, 1981). The data correspond to the viral isolates used in the construction of the evolutionary tree (Fig. 4). The line with an minimum square error was determined and represented. The point showing minimum fitness with the line corresponds to the NEB72 isolate.

Based on nucleotide sequencing data (Fig. 2), an evolutionary tree has been proposed which provides a relationship among 13 PRCV and TGEV isolates (Fig. 4). Since we are dealing with a limited number of isolates from each continent, it is understood that the inclusion of additional sequences from isolates of other areas (i.e., Japan) could show that certain lateral branches may become main branches. Only one isolate (NEB72) was out of place. According to the evolutionary tree, it should have been isolated at the same time as the PUR46 strain, since it has accumulated a similar number of point mutations and has the same 6-nt deletion which is present in the PUR46 strains. NEB72 probably represents a virus reintroduction, as the ones described in other viral systems (Beck and Strohmaier, 1987; Carrillo *et al.*, 1990). Least squares estimation of the origin of TGEV related coronaviruses demonstrates a significant constancy in the fixation of mutations with time, that is, the existence of a well-defined molecular clock (Kimura, 1983). The mutation fixation rate is of 0.95 ± 0.05 substitutions per year. As this rate was measured for 1260 nt, it can be expressed as $7 \pm 2 \times 10^{-4}$ substitutions per nucleotide and per year. This rate falls in the range reported for other RNA viruses (Domingo and Holland, 1988). The direction defined for the evolutionary process from the predicted origin supports the occurrence of two deletions: one of 6 nt in the lineage from the root to PUR46 strains and another of 672 nt in the lineage leading from TGEV to PRCVs. It may be concluded that the European PRCVs have been derived by a 672-nt dele-

tion from an enteric TGEV, since we have examined isolates preceding the PRCVs. In contrast, it cannot be guaranteed that the PUR46 emerged by a 6-nt deletion from an unknown ancestor. An alternative explanation could be that the other enteric isolates shown in Fig. 4 could have been derived from PUR46 by the addition of 6 nt.

It is interesting to note that the area deleted in the TGEV S gene to form the PRCVs contain repeated tetrameric (TTCC) or heptameric (AGTTTCC) sequences. These repeated sequences could be involved in inter- or intramolecular recombinations, by a copy choice mechanism (Lai, 1992), which could have originated the deletion. In coronaviruses and other RNA viruses containing positive-strand RNA genomes, recombinant clones have been isolated with borders at the crossover sites containing some sequence similarity (Banner and Lai, 1991; Raffo and Dawson, 1991; Cascone *et al.,* 1990). Since the putative crossover observed in the generation of PRCVs does not happen at homologous sequences, the deletion might have been originated by nonhomologous recombination. This mechanism has been previously involved in the evolution of coronaviruses, Sindbis virus, and plant viruses (Banner and Lai, 1991; Monroe and Schlesinger, 1983; Bujarski and Dzianott, 1991). If recombination has been the cause of the deletion present in the PUR46 and PRCVs, then two mechanisms of evolution would be involved in the antigenic variation of TGEV, point mutations and recombination.

ACKNOWLEDGMENTS

We are grateful to J. A. García for critical comments on the manuscript, to J. Palacín for his excellent technical assistance. F.G., and C.S. received fellowships from the Spanish Ministry of Education and Science. This investigation has been founded by grants from the Comisión Interministerial de Ciencia y Tecnología (Projects BIO 0214 and BIO 89-0668-603-03), European Communities (Project BAP 0464 E), NATO (Project CRG 900430), and Fundación Ramón Areces.

REFERENCES

ANSORGE, W. (1985). Fast and sensitive detection of protein and DNA bands by treatment with potassium permanganate. *J. Biochem. Biophys. Methods* **11,** 13–20.

BANNER, L. R., and LAI, M. M. C. (1991). Random nature of coronavirus RNA recombination in the absence of selection pressure. *Virology* **185,** 441–445.

BECK, E., and STROHMAIER, K. (1987). Subtyping of European foot-and-mouth disease virus strains by nucleotide sequence determination. *J. Virol.* **61,** 1621–1629.

BOHL, E. H., GUPTA, R. K. P., OLQUÍN, M. V. F., and SAIF, L. (1972). Antibody responses in serum, colostrum and milk of swine after infection or vaccination with transmissible gastroenteritis virus. *Infect. Immun.* **6,** 289–301.

BRITTON, P., and PAGE, K. W. (1990). Sequence of the S gene from a virulent British field isolate of transmissible gastroenteritis virus. *Virus Res.* **18,** 71–80.

BRITTON, P., PAGE, K. W., MAWDITT, K., and POCOCK, D. H. (1990). "Sequence Comparison of Porcine Transmissible Gastroenteritis Virus (TGEV) with Porcine Respiratory Coronavirus," Seventh International Congress of Virology, pp. P6–O18. IUMS, Berlin.

BROWN, J., and CARTWRIGHT, S. F. (1986). New porcine coronavirus? *Vet. Rec.* **119,** 282–283.

BUJARSKI, J. J., and DZIANOTT, A. M. (1991). Generation and analysis of nonhomologous RNA-RNA recombinants in Brome mosaic virus: Sequence complementarities at crossover sites. *J. Virol.* **65,** 4153–4159.

CALLEBAUT, P., CORREA, I., PENSAERT, M., JIMÉNEZ, G., and ENJUANES, L. (1988). Antigenic differentiation between transmissible gastroenteritis virus of swine and a related porcine respiratory coronavirus. *J. Gen. Virol.* **69,** 1725–1730.

CARRILLO, C., DOPAZO, J., MOYA, A., GONZÁLEZ, N., MARTINEZ, M. A., SAIZ, J. C., and SOBRINO, F. (1990). Comparison of vaccine strains and the virus causing the 1987 foot-and-mouth disease outbreak in Spain: epizootiological analysis. *Virus Res.* **15,** 45–56.

CASCONE, P. J., CARPENTER, C. D., LI, X. H., and SIMON, A. E. (1990). Recombination between satellite RNAs of turnip crinkle virus. *EMBO J.* **9,** 1709–1715.

CORREA, I. I., GEBAUER, F., BULLIDO, M. J., SUÑÉ, C., BAAY, M. F. D., ZWAAGSTRA, K. A., POSTHUMUS, W. P. A., LENSTRA, J. A., and ENJUANES, L. (1990). Localization of antigenic sites of the E2 glycoprotein of transmissible gastroenteritis coronavirus. *J. Gen. Virol.* **71,** 271–279.

COX, E., HOOYBERGHS, J., and PENSAERT, M. B. (1990a). Sites of replication of a porcine respiratory coronavirus related to transmissible gastroenteritis virus. *Res. Vet. Sci.* **48,** 165–169.

COX, E., PENSAERT, M. B., CALLEBAUT, P., and VAN DEUN, K. (1990b). Intestinal replication of a porcine respiratory coronavirus closely related antigenically to the enteric transmissible gastroenteritis virus. *Vet. Microbiol.* **23,** 237–243.

DE GROOT, R. J., LUYTJES, W., HORZINEK, M. C., VAN DER ZEIJST, B. A. M., SPAAN, W. J. M., and LENSTRA, J. A. (1987). Evidence for a coiled-coil structure in the spike protein of coronaviruses. *J. Mol. Biol.* **196,** 963–966.

DELMAS, B., RASSCHAERT, D., GODET, M., GELFI, J., and LAUDE, H. (1990). Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike protein. *J. Gen. Virol.* **71,** 1313–1323.

DOMINGO, E., and HOLLAND, J. J. (1988). High error rates, population equilibrium and evolution of RNA replication systems. *In* "RNA Genetics" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), Vol. 3, pp. 3–36. CRC Press, Boca Ratón, FL.

DOYLE, L. P., and HUTCHINGS, L. M. (1946). A transmissible gastroenteritis in pigs. *J. Am. Vet. Med. Assoc.* **108,** 257–259.

DURET, C., BRUN, A., GUILMOTO, H., and DAUVERGNE, M. (1988). Isolement, identification et pouvoir pathogène chez le porc d'un coronavirus apparenté au virus de la gastro-entérite transmissible. *Rec. Méd. Vét.* **164,** 221–226.

EFRON, B. (1982). "The Jackknife, the Bootstrap and Other Resampling Plans." Society for Industrial and Applied Mathematics, Philadelphia.

ENJUANES, L., and VAN DER ZEIJST, B. A. M. (1992). Molecular basis of transmissible gastroenteritis coronavirus (TGEV) epidemiology. *In* "Coronaviruses" (S. G. Siddell, Ed.), Plenum, New York.

FELSENSTEIN, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39,** 783–791.

FELSENSTEIN, J. (1990). "PHYLIP Manual Version 3.3." Editors. University Herbarium. University of California, Berkeley, California.

FICHOT, O., and GIRARD, M. (1990). An improved method for sequencing of RNA templates. *Nucleic Acids Res.* **18,** 6162.

FITCH, W. M., and MARGOLIASH, E. (1967). Construction of phylogenetic trees. *Science* **155,** 279–284.

FURUUCHI, S., SHIMIZU, Y., and KUMAGAI, T. (1976). Vaccination of pigs with an attenuated strain of transmissible gastroenteritis virus. *Am. J. Vet. Res.* **37,** 1401–1404.

GARWES, D. J., LUCAS, M. H., HIGGINS, D. A., PIKE, B. V., and CARTWRIGHT, S. F. (1978). Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* **3,** 179–190.

GARWES, D. J., STEWART, F., CARTWRIGHT, S. F., and BROWN, I. (1988). Differentiation of porcine coronavirus from transmissible gastroenteritis virus. *Vet. Rec.* **122,** 86–87.

GEBAUER, F., POSTHUMUS, W. A. P., CORREA, I., SUÑÉ, C., SÁNCHEZ, C. M., SMERDOU, C., LENSTRA, J. A., MELOEN, R., and ENJUANES, L. (1991). Residues involved in the formation of the antigenic sites of the S protein of transmissible gastroenteritis coronavirus. *Virology* **183,** 225–238.

JACOBS, L., DE GROOT, R., VAN DER ZEIJST, B. A. M., HORZINEK, M. C., and SPAAN, W. (1987). The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): Comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV). *Vir. Res.* **8,** 363–371.

JUKES, T. H., and CANTOR, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132. Academic Press, New York.

KIMURA, M. (1983). "The Neutral Theory of Molecular Evolution." Cambridge Univ. Press, London.

KING, A. M. Q. (1988). Genetic recombination in positive strand RNA viruses. *In* "RNA Genetics" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), Vol. 2, pp. 149–165. CRC Press. Boca Raton, FL.

LAI, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56,** 61–79.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685.

LEVINE, A. J. (1984). Viruses and differentiation: The molecular basis of viral tissue tropisms. *In* "Concepts in Viral Pathogenesis" (A. L. Notkins and M. B. A. Oldstone, Eds.), pp. 130–134. Springer-Verlag, New York.

MCCLURKIN, A. W., and NORMAN, J. O. (1966). Studies on transmissible gastroenteritis of swine. II. Selected characteristics of a cytopathogenic virus common to five isolates of transmissible gastroenteritis. *Can. J. Comp. Med. Vet. Sci.* **30,** 190–198.

MONROE, S. S., and SCHLESINGER, S. (1983). RNAs from two independently isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their 5′ ends. *Proc. Natl. Acad. Sci. USA* **80,** 3279–3283.

PENSAERT, M., CALLEBAUT, P., and VERGOTE, J. (1986). Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Vet. Q.* **8,** 257–260.

RAFFO, A. J., and DAWSON, W. O. (1991). Construction of Tobacco mosaic virus subgenomic replicons that are replicated and spread systemically in tobacco plants. *Virology* **184,** 277–289.

RASSCHAERT, D., and LAUDE, H. (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* **68,** 1883–1890.

RASSCHAERT, D., DUARTE, M., and LAUDE, H. (1990). Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. *J. Gen. Virol.* **71,** 2599–2607.

SAITOU, N. M., and NEI, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

SÁNCHEZ, C. M., JIMÉNEZ, G., LAVIADA, M. D., CORREA, I., SUÑÉ, C., MARÍA, J. B., GEBAUER, F., SMERDOU, C., CALLEBAUT, P., ESCRIBANO, J. M., and ENJUANES, L. (1990). Antigenic homology among coronaviruses related to transmissible gastroenteritis virus. *Virology* **174,** 410–417.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SIDDELL, S. G., WEGE, H., and TER MEULEN, V. (1982). The structure and replication of coronaviruses. *Curr. Topics Microbiol. Immunol.* **99,** 131–163.

SOKAL, R. R., and ROHLF, F. J. (1981). "Biometry." Freeman, New York.

SOURDIS, J., and NEI, M. (1988). Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* **45,** 298–311.

SPAAN, W., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69,** 2939–2952.

SUÑÉ, C., JIMÉNEZ, G., CORREA, I., BULLIDO, M. J., GEBAUER, F., SMERDOU, C., and ENJUANES, L. (1990). Mechanisms of transmissible gastroenteritis coronavirus neutralization. *Virology* **177,** 559–569.

UNDERDAHL, N. R., MEBUS, C. A., STAIR, E. L., RHODES, M. B., MCGILL, L. D., and TWIEHAUS, M. J. (1974). Isolation of transmissible gastroenteritis virus from lungs of market-weight swine. *Am. J. Vet. Res.* **35,** 1209–1216.

WESLEY, R. D. (1990). Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequence of the upstream polymerase gene of transmissible gastroenteritis virus (Miller strain). *Adv. Exp. Med. Biol.* **276,** 301–306.

WESLEY, R. D., WOODS, R. D., and CHEUNG, A. K. (1990a). Genetic basis for the pathogenesis of transmissible gastroenteritis virus. *J. Virol.* **64,** 4761–4766.

WESLEY, R. D., WOODS, R. D., HILL, H. T., and BIWER, J. D. (1990b). Evidence for a porcine respiratory coronavirus, antigenically similar to transmissible gastroenteritis virus, in the United States. *J. Vet. Diagn. Invest.* **2,** 312–317.

WESLEY, R. D., WOODS, R. D., and CHEUNG, A. K. (1991). Genetic analysis of porcine respiratory coronavirus, an attenuated variant of transmissible gastroenteritis virus. *J. Virol.* **65,** 3369–3373.