Exp.Rep AHRI.,No.40：90~110(2004)

Phylogenetic analysis of classical swine fever virus in Taiwan

C. H. Pan[1,2], **M. H. Jong**[1], **T. S. Huang**[1], **H. F. Liu**[3], **S. Y. Lin**[1], and **S. S. Lai**[2]

[1]Department of Hog Cholera, Animal Health Research Institute, Council of Agriculture, Taipei, Taiwan
[2]Department of Veterinary Medicine, National Taiwan University, Taipei, Taiwan
[3]Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan

Received July 22, 2004; accepted December 10, 2004
Published online February 10, 2005

Summary. Two envelope glycoprotein (E^{rns} and E2) regions of the classical swine fever virus (CSFV) were amplified by RT-PCR and sequenced directly from 158 specimens collected between 1989 and 2003 in Taiwan. Phylogenetic analysis of the two regions revealed a similar tree topology and the E^{rns} region provided better discrimination than the E2 region. One hundred and fifteen isolates out of the 158 isolates were clustered within subgroup 2.1 (further classified as 2.1a and 2.1b) and 2.2, which were considered to be likely of the introduced strains, whereas the remaining 43 isolates were clustered within subgroup 3.4 and were considered to be of the endemic strains. The subgroup 2.1a viruses were first detected in 1994 and predominated from 1995 onwards. However, subgroup 3.4 viruses were prevalent in the early years, not being isolated after 1996. We have observed a dramatic switch in genotype from subgroup 3.4 to 2.1a. The subgroup 2.1a isolates are closely related to the Paderborn and Lao isolates, whereas 2.1b isolates have a close relationship to the Chinese Guangxi isolates. The phylogenetic tree of 27 CSFV sequences based on the complete envelope glycoprotein gene (E^{rns}–E2) displayed better resolution than that based on the complete open reading frame.

Introduction

Classical swine fever (CSF), previously referred to as hog cholera, is a significant infectious disease of swine caused by classical swine fever virus (CSFV). The causative agent of CSFV belongs to the genus *Pestivirus* within the family *Flaviviridae* [21]. The genome of CSFV consists of a positive-stranded RNA approximately 12.3 kb in length, and comprises a single large open reading frame (ORF) spanning 11,694 nucleotides that encode a polyprotein composed of 3,898 amino acids [20]. The polyprotein is proteolytically processed to produce both

*Corresponding Author
Roprintod from Archives of Virology(2005)150:1101~1119

structural and nonstructural polypeptides [25]. The viral proteins are arranged in the following order (from the N to the C terminal): N^{pro}, C, E^{rns}, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B. The structural proteins represented by the capsid protein, C and three envelope glycoproteins, E^{rns}, E1 and E2, while the remaining proteins are presumably nonstructural proteins [21]. Three glycoproteins are associated with the envelope of the virion [32].

Phylogenetic analysis of different regions of the viral genome pertaining to the molecular discrimination of CSFV isolates has previously been reported. The regions analyzed include the 5′NTR [6, 7, 10, 18, 27, 28], E2 [18, 19, 33], NS5B [3], 5′NTR and E2 [1, 8], 5′NTR, E2 and NS5B [18, 24], 5′NTR and NS5B [9], E1/E2 [17, 37], E2 and NS5B [34, 35], 5′NTR, E2/NS2 and NS2 [29], 3′NTR [2, 36], and the complete open reading frame (ORF) [22]. Three regions of the CSFV genome (5′NTR, E2 and NS5B) have been widely sequenced and used in genetic analyses to investigate the diversity among isolates. A tentative assignment of world isolates of CSFV by genotyping has been divided it into three groups with three or four subgroups: 1.1, 1.2, 1.3; 2.1, 2.2, 2.3; 3.1, 3.2, 3.3, 3.4 [24].

Molecular epidemiological analyses of CSFV isolates distributed in the world are a useful to trace the geographic origins of CSFV and to understand how they spread. A large epidemic of CSF outbreaks occurred in the European Community at the beginning of 1997. The disease presumably spread to the Netherlands, Italy, Spain, and eventually to Belgium [8]. Sequence analysis based on 5′NTR and E1/E2 regions revealed that the isolates of the Dutch outbreaks of 1997–1998 were closely linked to a small CSF outbreak at Paderborn, Germany in 1996 [37]. Analysis of the E2 and partial 5′NTR sequence data showed that the Paderborn isolate (subgroup 2.1) triggered the Dutch outbreak [22].

Records of CSF in Taiwan date back to 1938 during occupancy by Japan. In an effort to control this highly contagious disease, the live attenuated lapinized hog cholera vaccine strain (LPC) has been widely used in the field since 1958. Vaccination significantly decreased the rate of occurrence. However sporadic outbreaks were still reported every year. The aim of this work was to analyze the phylogenetic relationship of CSFV isolates collected in Taiwan between 1989 and 2003.

Materials and methods

Virus samples

CSFV were isolated from the sick or dead pigs collected subsequently from the CSF suspected farms, which widely distributed on the island of Taiwan between 1989 and 2003. Each virus was isolated from an individual pig farm and confirmed by immuno-FITC staining. Totally 158 isolates were obtained and used for this study. Tissue specimens of tonsils, spleens and lymph nodes were prepared as 10% (w/v) suspensions in MEM medium (Eagles). The LPC/AHRI vaccine strain was taken from the spleens and lymph nodes of inoculated rabbits that were routinely made for commercial vaccine production at the Animal Health Research Institute (AHRI) at Tamsui, Taiwan.

RT-PCR and sequencing

Using the nucleotide sequences of 20 currently available complete sequences in GenBank, we aligned the regions of structural protein genes using the program MegAlign (clustal W

method) from the Lasergene program package (DNASTAR, 2001). RT-PCR primers were designed from highly conserved regions, reflected especially in the 3′ end portion of the primers in an effort to amplify different genotypes of CSFV. Seven sets of primers located in the C, E^{rns}, E1, E2 and P7 gene regions were chosen for RT-PCR and DNA sequence analyses. The sequences, locations and amplified regions of these primers are summarized in Table 2.

Viral RNA was extracted directly from 100 µl of the 10% (w/v) suspension clinical tissue samples using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. The dried RNA pellets were resuspended in 100 µl of diethyl pyrocarbonate (DEPC)-treated water for immediate use or stored at −70 °C. Reverse transcription (RT) and subsequent PCR were performed in one tube with a single reaction buffer. The protocol used was an adaptation of the two-step RT-PCR assay described by Liu et al. [16] The RT-PCR reaction was carried out in a total volume of 50 µl. The mixtures contained 1 U Supertherm DNA Polymerase (JMR, UK), buffer supplied by the manufacturer (containing 1.5 mM magnesium chloride), 8 U of RNase inhibitor (Promega), 2 U of AMV reverse transcriptase (Promega), 10 pmoles each of the sense and antisense primers (described in Table 1), 0.1 mM of each dNTP and 5 µl of RNA samples. The single-step RT-PCR reaction was carried out in a Perkin Elmer Gene Amp 9600 thermal cycler using the following programme: 40 min at 42 °C for RT, denaturation for 1 min at 94 °C, followed by PCR with 35 cycles of denaturation for 50 s at 94 °C, annealing for 1 min at 50 °C, extension for 1 min at 72 °C and a final extension step at 72 °C for 7 min. This programme was used in reactions for each of the seven primer sets. RT-PCR products were analyzed by electrophoresis through 1.5% agarose gels containing 0.5 mg/ml ethidium

Table 1. Isolation years and genotypes of 158 Taiwanese CSFV isolates sequentially listed according to period isolated

Isolates	Isolation year	Genotypes
78-KS	1989	3.4
79-60	1990	3.4
82-17	1993	3.4
82-18	1993	3.4
82-19	1993	3.4
82-34	1993	3.4
82-35	1993	3.4
82-38	1993	3.4
82-39	1993	3.4
82-40	1993	3.4
P97	1993	3.4
83-17	1994	3.4
83-18	1994	3.4
83-19	1994	3.4
83-49	1994	3.4
83-51	1994	3.4
83-52	1994	3.4
83-55	1994	3.4
83-56	1994	3.4
83-57	1994	3.4
83-58	1994	3.4

(*continued*)

Table 1 (*continued*)

Isolates	Isolation year	Genotypes
83-s59	1994	3.4
94.4	1994	3.4
83-60	1994	3.4
83-s68	1994	3.4
83-101	1994	3.4
83-s106	1994	2.1a
83-114	1994	3.4
83-118	1994	2.1a
83-129	1994	2.1a
83-130	1994	2.1a
83-182	1994	3.4
83-PT	1994	3.4
84-ES1	1995	2.1a
84-ES4	1995	2.1a
84-S2	1995	2.1a
84-HC1	1995	2.1a
84-YL1	1995	2.1a
84-YL2	1995	2.1a
84-FL1	1995	2.1a
84-FL2	1995	2.1a
84-HL	1995	2.1a
84-s10	1995	2.1a
84-s11	1995	2.1a
84-s12	1995	2.1a
84-16	1995	3.4
84-TP	1995	2.1a
84-19	1995	3.4
84-20	1995	3.4
84-25	1995	2.1a
84-30	1995	2.1a
84-31	1995	2.1a
84-32	1995	2.1a
84-36	1995	2.1a
84-55	1995	3.4
84-56	1995	2.1a
84-89	1995	3.4
84-98	1995	2.1a
84-99	1995	2.1a
84-100	1995	2.1a
84-104	1995	3.4
84-105	1995	2.1a
84-s107	1995	2.1a
84-s108	1995	2.1a
84-KS1	1995	2.2
84-109	1995	2.1a

(*continued*)

Table 1 (*continued*)

Isolates	Isolation year	Genotypes
84-C	1995	2.2
84-D	1995	2.2
84-E	1995	2.2
84-P	1995	3.4
84-Q	1995	3.4
84-R	1995	3.4
Q84-121	1995	2.1a
Q84-122	1995	2.1a
84-110	1995	2.1a
84-112	1995	2.1a
84-s113	1995	2.1a
84-117	1995	2.1a
84-118	1995	2.1a
84-120	1995	2.1a
84-121	1995	3.4
84-124	1995	2.1a
84-127	1995	2.1a
84-131	1995	3.4
84-132	1995	2.1a
84-134	1995	2.1a
84-140	1995	3.4
84-141	1995	2.1a
Q84-348	1995	2.1a
Q84-349	1995	2.1a
D868	1996	3.4
85-DU1	1996	2.1a
85-DU	1996	2.1a
85-10	1996	2.1a
85-12A	1996	3.4
85-101	1996	2.1a
85-176	1996	2.1a
85-EL	1996	2.1a
Q85-17	1996	2.1a
Q85-91	1996	2.1a
Q85-116	1996	2.1a
Q85-351	1996	2.1a
85-TN264	1996	2.1a
9654	1996	2.1a
9664	1996	2.1a
85-CF	1996	2.1a
85-TD2	1996	2.1a
85-TD3	1996	2.1a
86-HL1	1997	2.1a
86-HL2	1997	2.1a
86-HL3	1997	2.1a

(*continued*)

Table 1 (*continued*)

Isolates	Isolation year	Genotypes
Q87-278	1998	2.1a
88-PT	1999	2.1a
88-CH	1999	2.1a
88-CC25	1999	2.1a
89-YL1	2000	2.1a
Q89-240	2000	2.1a
Q89-273	2000	2.1a
89-SC	2000	2.1a
90-IL	2001	2.1a
90-R0143	2001	2.1a
90-KS	2001	2.1a
90-YL1	2001	2.1b
90-YL2	2001	2.1b
90-YL3	2001	2.1b
90-YL5	2001	2.1a
90-TP1	2001	2.1a
90-CH1	2001	2.1b
90-KS	2001	2.1a
90-CH14	2001	2.1a
90-CH15	2001	2.1a
90-CH23	2001	2.1a
90-TD1	2001	2.1a
90-PT50	2001	2.1a
90-TN1	2001	2.1b
90-TN2	2001	2.1b
90-TN3	2001	2.1b
90-PF	2001	2.1a
90-LSH1	2001	2.1a
90-TD2	2001	2.1a
Q90-48	2001	2.1a
Q90-152	2001	2.1a
90-SC2	2001	2.1a
90-LSH2	2001	2.1a
90-YL11	2001	2.1a
Q90-278	2001	2.1a
90-CY	2001	2.1a
91-NT1	2002	2.1a
Q91-52	2002	2.1a
Q91-84	2002	2.1b
Q91-117	2002	2.1a
Q91-178	2002	2.1a
91-ST	2002	2.1a
Q92-39	2003	2.1a
Q92-43	2003	2.1a
92-TN1	2003	2.1a
92-TC1	2003	2.1a
92-CS1	2003	2.1a

Table 2. Oligonucleotide primers used for RT-PCR and sequencing

Primer	Sequence	Position[a]	Region	Length
FE-1F (sense)	5′ TTAAARATAGCCCCAAAAGAGCATG 3′	937–961	C	419 bp
FE-1R (antisense)	5′ CTGGCGTCCATCATYCCGYGTAT 3′	1355–1333	E^{rns}	
E0-2F (sense)[b]	5′ CAGCAAGCYATGTAYCTTAGAGGG 3′	1222–1245	E^{rns}	542 bp
E0-2R (antisense)[b]	5′ GCACCTTGYCTGGCATTCTCTAT 3′	1763–1741	E^{rns}	
FE-3F (sense)	5′ ACYCTGACYGGCTGCAAGAAAGG 3′	1573–1595	E^{rns}	437 bp
FE-3R (antisense)	5′ CCCCCCATYTCATGRAGAATCTT 3′	2009–1987	E1	
FE-4F (sense)	5′ CTCCCCAAAAAYACAAARATAATAGG 3′	1927–1952	E1	509 bp
FE-4R (antisense)	5′ TGTGCCCCGGTYACCAGYAACAGCCA 3′	2435–2410	E1	
E2-5F (sense)[c]	5′ CAACCACGGCATTCCTCATYTG 3′	2345–2366	E1	504 bp
E2-5R (antisense)[c]	5′ TGACACCCGTCCACCCTATTG 3′	2848–2828	E2	
FE-6F (sense)	5′ CCTGTGGTCAARGGGAAGTACAA 3′	2764–2786	E2	380 bp
FE-6R (antisense)	5′ ACTCTGTAACCCGTCTCATTTGC 3′	3143–3121	E2	
FE-7F (sense)	5′ TGCAGGTGGTGCGGYTTYGACTT 3′	3049–3071	E2	563 bp
FE-7R (antisense)	5′ GTGTGAGTRATTAAGTTCCCTAT 3′	3611–3589	P7	

[a]Positions correspond to the sequence of the Alfort Tubingen strain
[b]E0-2F, 2R for the phylogenetic analysis of the E^{rns} tree
[c]E2-5F, 5R for the phylogenetic analysis of the E2 tree

bromide. Two sets of primers (E0-2F, 2R and E2-5F, 5R) were used to amplify by RT-PCR and directly sequence 158 clinical specimens collected between 1989 and 2003 (Tables 1 and 2).

All 353 RT-PCR products were purified using the QIAQuick Purification Kit (QIAGEN) without gel extraction. DNA sequences of the RT-PCR products were determined by the direct sequencing method using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) with the original RT-PCR primers. DNA fragments were sequenced in both directions using the sense and antisense primers. The cycle sequencing products were analyzed on an ABI 3100 genetic analyzer (Applied Biosystems).

Seven sets of primers (see Table 2) added individually to each of seven tubes were used to amplify seven viruses belonging to five distinct genotypes. These represented subgroup 2.1a (83-s106, 92-TC1), subgroup 2.1b (90-YL1), subgroup 2.2 (84-KS1), subgroup 3.4 (79-60, 85-12A) and subgroup 1.1 (LPC/AHRI). Following RT-PCR, nucleotide sequencing was performed using the original primers. Employing the sequence Navigator software (version 1, PE Applied Biosystems), seven fragments were adjoined to yield a sequence 2,385 bp in length for phylogenetic analysis of the complete envelope glycoprotein gene (E^{rns}–E2 region).

Phylogenetic trees

Nucleotide sequence alignments were carried out using the Clustal W program (DNASTAR, 2001). Two fragments, 431 and 190 bp, corresponding to 1269–1699 (E^{rns}) and 2517–2706 (E2) of the Alfort Tubingen sequence [20], respectively, were chosen for the phylogenetic analysis. For comparison, 27 reference strain sequences representing three groups and six subgroups were obtained from the GenBank and analyzed with Taiwanese CSFV isolates. They were Shimen (AF092448), cF-114 (AF333000), HCLV (AF091507), Riems (U45477), Chinese C strain (Z46258), Brescia (M31768), Brescia (AF091661), ALD (D49532), GPE^{-}

(D49533), CAP (X96550), KC (AF099102), Glentorf (U45478), Eystrup (AF326963), Alfort A19 (U90951), Alfort 187 (X87939), Paderbron (AY072924), 4/sw/NL/Venhorst (AF084050), L119 (AY283667), s7D2 (L36171), GXWZ02 (AY367767), CW2002 (AF521712), Alfort Tubingen (J04358), c1W (L36164), s7D (L36170), 39 (AF407339), L67 (AY283663) and P97 (L49347). The GenBank accession numbers are noted in parentheses. Empirical transition/transversion rations of 6.30 and 5.12 for the E^{rns} and E2 genes, respectively, were estimated using the Puzzle 4.0 program [30, 31] and were used with Kimura's two-parameter model [12] to calculate evolutionary distances. The unrooted neighbour-joining phylogenetic trees were constructed using the Phylip software package [4] with the neighbor-joining (NJ) method [26]. Bootstrap analyses were performed by 1000 resamplings of the data sets. Bootstrap values ≥70% were considered to be statistically significant for the grouping [5]. Phylogenetic trees were visualized using the TreeView 1.6 program [23].

Phylogenetic analyses were also performed on 27 sequences including six domestic isolates, LPC/AHRI vaccine strain, reference strain GXWZ02 and 19 reference strains described by Oleksiewicz et al. in 2003, based on the envelope glycoprotein gene using NJ and the maximun likehood (ML) methods. Given that the ML is a method possessing a statistical evaluation of the branch length, bootstrapping was not performed.

Results

RT-PCR and nucleotide sequences of E^{rns} and E2 genes

Two sets of primers (E0-2F, E0-2R and E2-5F, E2-5R) designed from the E^{rns} and E2 regions of the viral genome were successfully used to amplify each of the 158 CSFV and generated clear bands (about 542 and 504 bp in length, respectively) of the amplicons as visualized in ethidium bromide-stained agarose gels (not shown). The sequences from the positive and the negative strands overlapped by at least 431 bp in the E^{rns} and 190 bp in the E2 regions, which were selected for phylogenetic analysis.

Phylogenetic analysis of E^{rns} and E2 genes

In an effort to improve the survey, identical sequences in the E^{rns} region were omitted. Eventually, 67 nucleotide sequences (chosen from 158 Taiwanese isolates), LPC/AHRI sequence and reference strain sequences were used for the phylogenetic analysis and carried with 1000 bootstrap resamplings based on the 190 nucleotides of the E2 fragment and 431 nucleotides of the E^{rns} fragment. Phylogenetic trees showed that four distinct CSFV genotypes existed in Taiwan between 1989 and 2003 and were classified into subgroups 2.1a, 2.1b, 2.2 and 3.4 (Fig. 1A and B). One hundred and fifteen out of the 158 isolates (72.7%) were clustered within group 2 and could be further divided into three subgroups (2.1a, 2.1b and 2.2). Subgroup 2.1a is the largest group containing 103 Taiwanese isolates (40 isolates are shown in the trees) isolated between 1994 and 2003 that clustered closely with reference strains comprising of the German isolate (Paderborn), Dutch isolate (Venhorst) and Lao isolate (L119). Subgroup 2.1b contains eight Taiwanese isolates (three isolates are shown in the trees) isolated between 2001 and 2002 that clustered with the Chinese isolate (GXWZ02) and Korean isolate (CW2002). Only four Taiwanese isolates (two isolates are shown

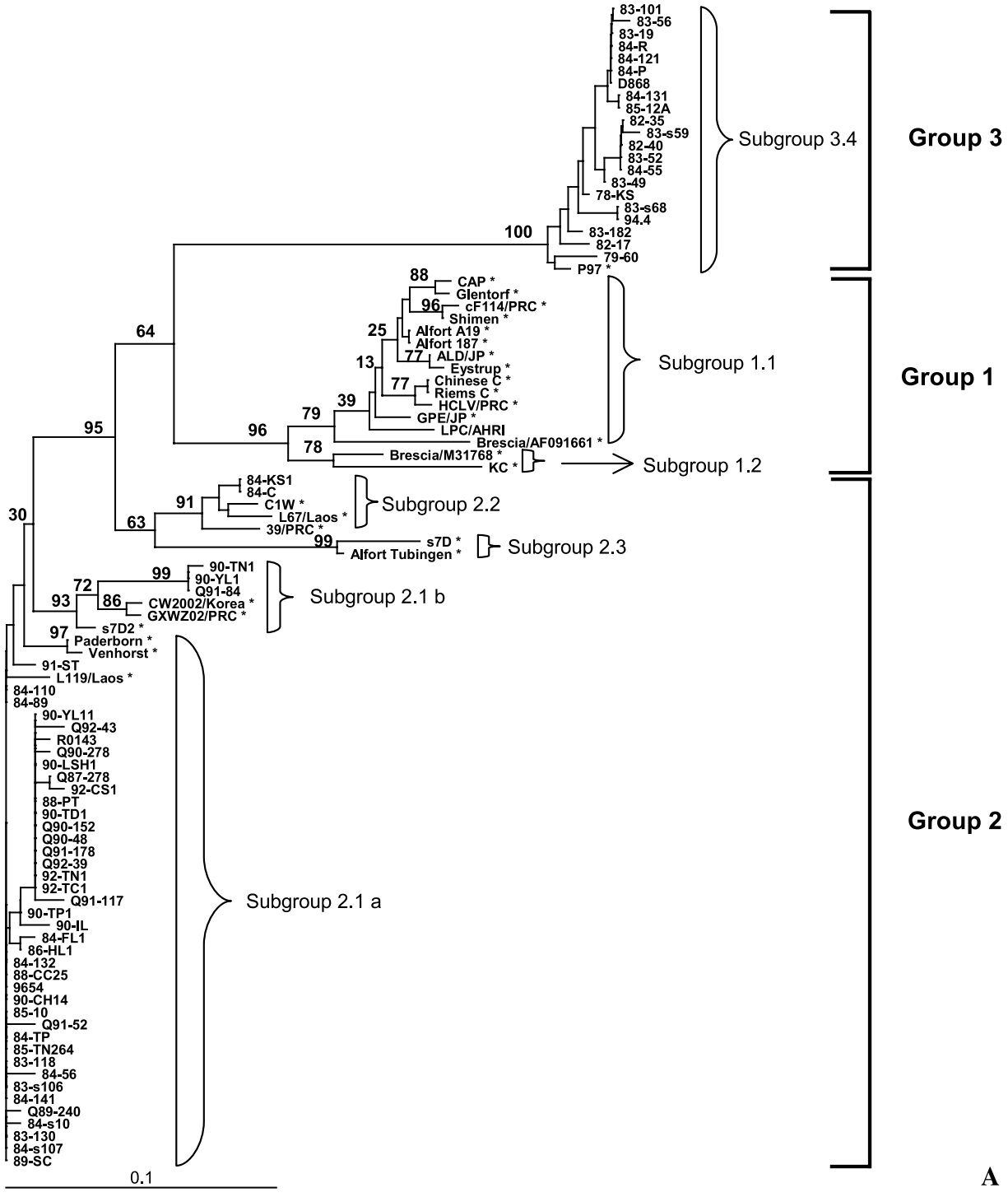


Fig. 1 (*continued*)

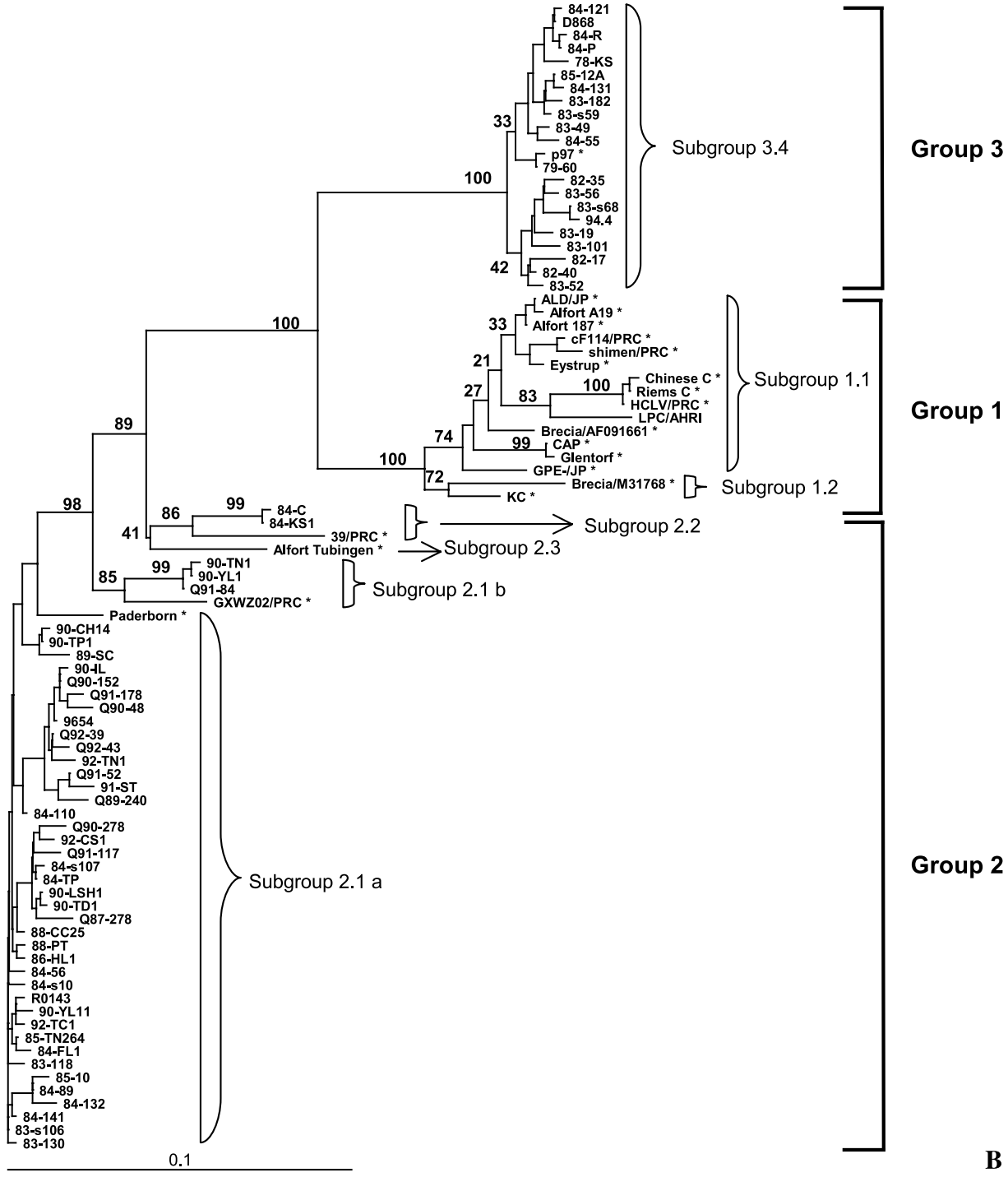


Fig. 1A, B. Unrooted phylogenetic trees of the 67 Taiwanese isolates, LPC/AHRI vaccine strain, and reference strains based on (**A**) 190 nucleotide fragment of the E2 gene (adding 27 reference strains) and (**B**) 431 nucleotide fragment of the E^{rns} gene (adding 20 reference strains). The trees were constructed by the neighbour-joining method. Evolutionary distances were calculated using Kimura's two-parameter model with a transition/transversion substitution ratio of 5.12 and 6.30, respectively. Bootstrap values (1000 bootstrap samples) are shown beside the branches as a percentage. Reference strains are marked with an asterisk (*)

in the trees) were included in subgroup 2.2. They were isolated in 1995 and clustered with reference strains comprising of c1W of Italy, L67 of Laos and 39 of China. None of the field viruses tested fell into subgroup 2.3, which was represented by reference strains comprising of Alfort Tubingen and s7D, or group 1 which was represented by 15 reference strains. The remaining 43 isolates (27.3%), including the Taiwanese reference strain P97, were clustered within subgroup 3.4 (22 isolates are shown in the trees). The LPC/AHRI vaccine strain was clustered in subgroup 1.1. All 158 Taiwanese isolates represented distinct genotypes are shown in Table 1.

Phylogenetic analysis of the two different genomic regions revealed similar groups and subgroups. In spite of the similar branching pattern, the E^{rns} region provided better discrimination than the E2 region, as supported by the higher bootstrap values. The bootstrap values for group 1, group 2 and subgroup 3.4 were all 100% in the E^{rns} tree (Fig. 1B) and 96%, 64% and 100%, respectively, in the E2 tree (Fig. 1A). Four vaccine strains of LPC/AHRI (Taiwan), HCLV (China), Riem C and Chinese C strain were clustered together in the E^{rns} tree but were not seen in the E2 tree (Fig. 1A and B). Chronological distribution of 158 Taiwanese isolates (Fig. 2) showed that three introduced virus strains, different in genotypes 2.1a, 2.1b and 2.2, were first detected in 1994, 2001 and 1995, respectively. The subgroup 2.1a viruses appeared in 1994 and resulted in a severe outbreak in 1995, then predominated in the field onwards. Subgroup 2.1b viruses were isolated only in 2001 and 2002. Subgroup 2.2 viruses were isolated only in 1995 and resulted in small outbreaks. Subgroup 3.4 viruses were mainly prevalent in Taiwan prior to 1996, and were not isolated from the field thereafter.

The greatest diversity in the 103 isolates of subgroup 2.1a, collected between 1994 and 2003, showed 3.1% in the E2 tree and 3.3% in the E^{rns} tree. The greatest diversity in the 43 isolates of subgroup 3.4, collected between 1989 and 1996, showed 3.3% in the E2 tree and 3.6% in the E^{rns} tree. The greatest diversity in

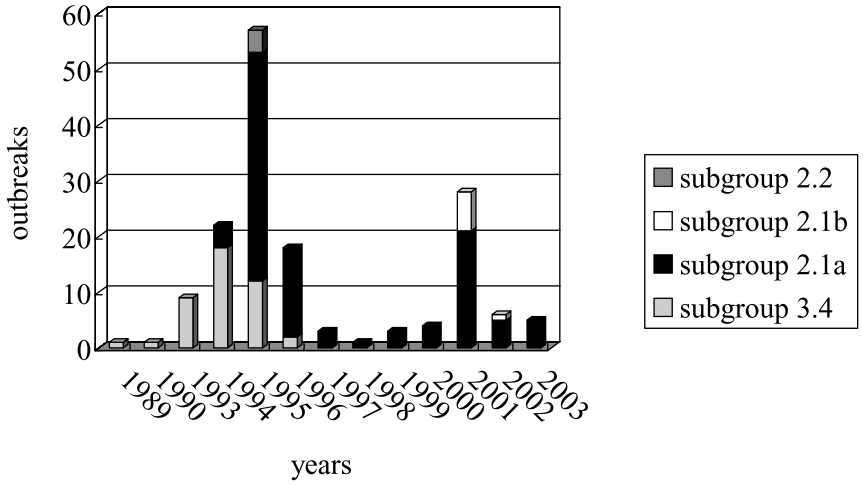
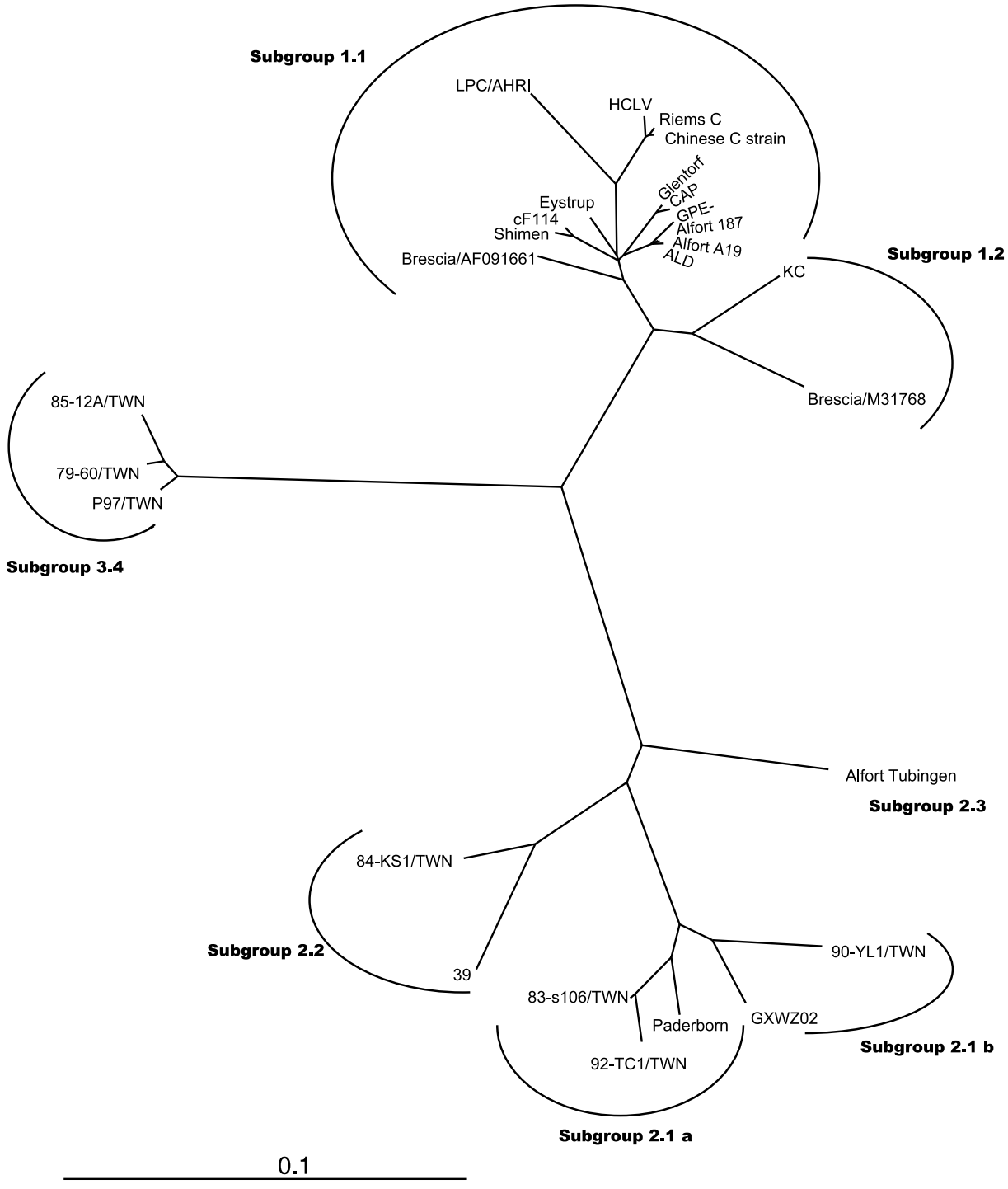


Fig. 2. The chronological and genotype distribution of 158 CSFV isolates in Taiwan between 1989 and 2003

the isolates of 2.1b, collected between 2001 and 2002, showed 0.5% in the E2 tree and 0.9% in the E^{rns} tree. The greatest diversity in subgroups 2.1a and 2.1b showed 6.8% in the E2 tree and 8% in the E^{rns} tree.



Phylogenetic analysis and sequence comparison of complete envelope protein gene (E^{rns}–E2)

In order to trace back the origin of the introduced strains and LPC vaccine strain, we sequenced the complete envelope glycoprotein gene (E^{rns}–E2 region) of six isolates and the LPC/AHRI vaccine strain. In an effort to cover the complete envelope glycoprotein genes consisting of 2,385 bp, seven RT-PCR reactions were performed in order to amplify the genotypically distinct viruses using seven sets of primers. Five different viral genotypes represented by subgroups 1.1, 2.1a, 2.1b, 2.2 and 3.4 were successfully amplified. Sequences of the seven viruses have been submitted to GenBank and have been assigned with accession numbers 92-TC1 (AY526726), 83-s106 (AY526727), 90-YL1 (AY526728), 84-KS1 (AY526729), 85-12A (AY526730), 79-60 (AY526731) and LPC/AHRI (AY526732). Phylogenetic analysis was performed on 27 sequences including the seven aforementioned sequences, 19 reference sequences described by Oleksiewicz et al. (2003) and the reference sequence GXWZ02 based on the complete envelope protein gene (E^{rns}–E2). The ML tree showed a similar topology as the ORF tree described by Oleksiewicz et al. [22]. The only difference was that the GPE^- was clustered together with the ALD and close to Alfort A19 and Alfort 187 (Fig. 3). Comparing the earliest isolate of 83-s106 (subgroup 2.1a) with the Paderborn strain showed 97.1% nucleotide and 98.5% amino acid identity based on the complete envelope glycoprotein gene sequence. In contrast, the 90-YL1 (subgroup 2.1b) and the Paderborn strain only showed 94.3% nucleotide and 97.4% amino acid identity. Comparing the LPC/AHRI vaccine sequence (subgroup 1.1) with the 83-s106 (2.1a), 90-YL1 (2.1b), 84-KS1 (2.2) and 79-60 (3.4) isolates showed 82.4%, 81.5%, 82.2%, 83.0% nucleotide and 89.8%, 89.3%, 89.8%, 89.7% amino acid identity, respectively, based on the complete envelope glycoprotein gene.

Discussion

Phylogenetic analysis based on the E^{rns} and E2 regions clustered Taiwanese field viruses into two major groups (2 and 3) containing three subgroups (2.1, 2.2 and

Fig. 3. Unrooted phylogenetic tree of 27 CSFV based on the complete envelope glycoprotein gene sequence (bases 1174–3558 from the Alfort Tubingen sequence), containing 2,385 nucleotides. The trees were constructed by the maximum likelihood method. The observed transition/transversion ratio in the sample sets was 6.20, and the observed nucleotide frequencies were A: 0.291, C: 0.216, G: 0.267, T: 0.226. All nodes in the tree were statistically significant ($P < 0.01$). The accession numbers of the sequences included in this tree are as follows: Alfort A19 (U90951), Alfort 187 (X87939), Alfort Tubingen (J04358), ALD (D49532), GPE^- (D49533), CAP (X96550), Glentorf (U45478), Eystrup (AF326963), Shimen (AF092448), cF114 (AF333000), HCLV (AF091507), Riems C (U45477), Chinese C strain (Z46258), Paderbron (AY072924), 39 (AF407339), GXWZ02 (AY367767), Brescia (M31768), Brescia (AF091661), KC (AF099102), p97/TWN (L49347), 92-TC1/TWN (AY526726), 83-s106/TWN (AY526727), 90-YL1/TWN (AY526728), 84-KS1/TWN (AY526729), 85-12A/TWN (AY526730), 79-60/TWN (AY526731) and LPC/AHRI (AY526732)

3.4; according to the nomenclature by Paton et al. [24]). Since the subgroup 2.1 could be further separated into two different clusters with high bootstrap values of 98% and 85% in the E^{rns} tree (Fig. 1B), we proposed that subgroup 2.1 should be further classified as 2.1a and 2.1b. Group 2 viruses (subgroups 2.1a, 2.1b and 2.2) showed a close relationship with the other viruses available in GenBank, and consequently we referred to them as the introduced strains. The subgroup 3.4 viruses existed in Taiwan prior to 1996, and as such, we referred to them as the endemic strains. In total, four distinct CSFV genotypes including one endemic genotype (subgroup 3.4) and three introduced genotypes (subgroup 2.1a, 2.1b and 2.2) existed in Taiwan between 1989 and 2003. None of the intertype strain was found from the field. Although three Taiwanese endemic strains (p97, 93.4 and 94.4) were classified into subgroup 3.4 by previous investigators [24], none of the Taiwanese introduced strains were referred to in that report and never in others. This is the first report to describe these new genotypes viruses in Taiwan. Epidemiological data showed that subgroup 2.1a viruses were introduced into Taiwan in 1994, caused outbreaks in 1995, and then predominated from 1995 onwards. However, the subgroup 3.4 viruses were mainly prevalent in Taiwan prior to 1996 and seemed to disappear from the field since it could not be isolated from the field thereafter (Fig. 2). We have observed a dramatic switch in genotype from subgroup 3.4 to 2.1a in 1996, which was not caused by genetic mutation of the endemic strains.

The subgroup 3.4 viruses containing Taiwanese isolates (p97, 94.4, 93.4) and Japanese isolates (Kanagawa/74, Okinawa/86) are some of the most distinctive viruses in the world and have only been isolated in Taiwan, island of Okinawa and Japan [27, 24]. The viruses belonging to this genotype have never been identified in China [33]. Epidemiological data showed that the subgroup 3.4 viruses possibly became ‘silent’ types in the Far East. Both subgroup 2.1a and subgroup 3.4 viruses belong to middle virulent strains [24]. The significance of this finding probably has its epidemiological links to the long-term application of the LPC vaccine in the field. A similar switch in genotype from Group 1 to Group 2 was also observed in China and Europe [24, 33]. It appeared that the Group 2 viruses caused a higher incidence of infection in the field. One interesting question is whether the switch in genotype from 3.4 to 2.1a affected the efficacy of the vaccine. Our vaccination and challenge tests showed that the LPC-immunized pigs were fully protected from challenge with 2.1a and 2.1b viruses (data unpublished). Similar results were also observed for the HCLV-immunized pigs that were fully protected from challenge with subgroup 2.1 and 2.2 viruses [33]. Results showed that the pigs immunized with lapinized vaccine strains (LPC or HCLV) were able to resist challenge even with genetic highly diverse CSFV isolates.

Phylogenetic analysis of the Taiwanese isolates of subgroup 2.1a with the known strains in GenBank indicated that it exhibited a closer relationship to the European and the Lao isolates including the German isolate (Paderborn) in 1996, the Dutch isolate (Venhorst) in 1997 and the Lao isolates (L119) in 2003. A determination of the origin of the Taiwanese 2.1a virus is an important issue in the molecular epidemiology of CSFV. The Paderborn isolate (subgroup 2.1) triggered

the Dutch outbreak, and then spread to Italy, Spain, and eventually to Belgium [8, 37]. The viruses of subgroup 2.1 that had been reported infrequently in Europe were no longer found after 1993. The CSFV isolates of the European epidemics in 1997–1998 and those that caused sporadic infection in different European countries before 1993 were clustered within subgroup 2.1 and were further clearly divided into two distinct clusters (E2 tree). Thus Greiser-Wilke et al. [8] postulated that the 1997–1998 epidemic outbreaks caused by a virus that had been newly introduced into Europe. However, given the lack of available sequence data in GenBank, the origin of the Paderborn isolate was still unknown. Although the Taiwanese subgroup 2.1a viruses possessed a close relationship to the Paderborn strain, they were introduced into Taiwan in 1994, well before 1996. Therefore a European origin seems unlikely. Thus we postulated that the Taiwanese subgroup 2.1a virus and the Paderborn strain originated from somewhere in Asia, and were introduced into Taiwan and Europe in 1994 and 1996, respectively. This hypothesis has been corroborated by the reference sequence of Lao strain (L119) (GenBank sequence data submitted in 2003), which is closely related to the Taiwanese subgroup 2.1a viruses in the E2 tree (lack of E^{rns} sequence data). From the geographical point of view, Laos and Taiwan are neighboring to China. So we believe that Lao and Taiwanese isolates may also have the same origin. It is generally suspected that the introduction of new subgroup viruses into Taiwan resulted from the smuggling of pigs and pig-related products.

In an effort to determine the optimal regions within the viral genome that would discriminate between CSFV isolates, Lowings et al. [18] compared three regions of CSFV comprised of E2 (190 bp), NS5B (210 bp) and 5′NTR (94 bp). It was shown that the E2 region provided the best discrimination, and that the NS5B region showed superior resolution to that of the 5′NTR region. Paton et al. [24] compared 55 isolates using three similar target regions comprised of E2 (190 bp), NS5B (409 bp) and 5′NTR (150 bp). It was determined that a grouping based on these three regions was very similar, whereas the statistical confidence separating the groups varied drastically. Therefore it was concluded that the most reliable classification of CSFV was obtained from the NS5B region. We believed that the difference was due to the shorter fragment length of the E2 region used for the phylogenetic analysis. The discriminative ability of the phylogenetic method depends on the regions of choice and the length of fragments analyzed. Higher resolution could be obtained by choosing a longer fragment. If the 190 bp region of E2 described by Paton et al. was expanded to 409 bp, the E2 region might have resulted in better discrimination than NS5B. Our analysis based on a 190 nucleotide region of E2 and a 431 nucleotide region of E^{rns} revealed a similar branching pattern, whereas the E^{rns} region provided better discrimination than the E2 region, as supported with higher bootstrap values. Subgroup 3.4 showed a tendency to be further divided into two clusters, and the four vaccine strains LPC/AHRI, HCLV, Riems C and Chinese C strain were clustered together in the E^{rns} tree, as supported by high bootstrap values but not seen in the E2 tree (Fig. 1A and B). In our opinion, the observed difference in results might be due to the length of the fragment analyzed. In order to investigate the effect of fragment

length on the outcome of the investigation, we used a short region comprising 190 bp within the 431 bp of the E^{rns} region to do the phylogenetic analysis. It was found that although the results still gave a similar tree topology, the bootstrap values decreased significantly, with one of the largest decrease being a reduction from 98% to 54% (data not shown). This suggested that the E2 gene should be expanded to the length at least 400 bp for a better bootstrap support.

Phylogenetic analysis could help to trace the origin of ancestor viruses. A known example is that GPE^- vaccine virus was obtained from an ALD strain after multiple passages in swine testicle, bovine testicle and guinea-pig kidney cells [11]. They were not clustered closely in both the E^{rns} and E2 trees. Previous studies including NS5B [3] and the polyprotein ORF phylogeny [22] showed the ALD and GPE^- vaccine viruses did not cluster together, even though a long fragment of the complete ORF sequence (containing 11,691 bp) was used for the analysis. Phylogenetic analysis of 20 sequences based on the complete polyprotein ORF resulted in some unexpected groupings; the GPE^- vaccine sequence was close to Alfort A19 and Alfort 187, while the ALD sequence clustered together with the Eystrup sequence [22]. In an effort to arrive at a more accurate discrimination of CSFV and to trace the origin of Taiwanese introduced strains, we constructed a maximum likelihood tree based on the nucleotide sequence of the complete envelope glycoprotein gene (containing 2,385 bp) using 27 sequences. The resulting tree showed that ALD was clustered together with the GPE^- sequence and was close to Alfort A19 and Alfort 187. The GPE^- had a longer branch than ALD (Fig. 3). The tree revealed that the GPE^- vaccine strain was derived from the progenitor of the ALD strain. Repeated analysis of the same 20 reference sequences described by Oleksiewicz et al. [22] but based on the complete envelope glycoprotein gene, showed that ALD still clustered with the GPE^- vaccine virus (data not shown). We therefore suggested that analysis based on the complete envelope glycoprotein gene provided better resolution. It was suitable for analyzing some of ambiguous sequences, which showed distinctive clustering in different regions like the ALD and GPE^- sequences, or the low bootstrap values in the trees. The length of the fragment and which regions in the genome should be targeted to afford a more accurate discrimination of virus strains during phylogenetic analysis is an important issue. An analysis of foot-and-mouth disease virus (FMDV) using the outer-capsid polypeptides (VP2, VP3 and VP1) resulted in a useful tree that discriminated 7 serotypes of FMDV [13]. Consequently, the polyprotein composed of the outer virion may serve as a favorable region for phylogenetic analyses in the cases of CSFV and FMDV.

We applied the phylogenetic trees to trace the origins of Lapinized vaccine strains. The LPC and HCLV vaccine strains were lapinized from the Rovac strain [14] and the Shimen strain [39], respectively, but these two vaccine strains were clustered together both in the E^{rns} (83% bootstrap value) and the complete envelope glycoprotein gene trees. Previous studies showed that one notable insertion of 12–13 continuous nucleotides is found in the 3′NTR of LPC, HCLV and the Chinese C strain [38, 39]. Our sequence data in the 3′NTR showed that the LPC/AHRI vaccine strain also possessed the same insertion of 12 continuous

nucleotides (data not shown). Therefore we predicted that the LPC and HCLV strains shared a common ancestor. However, LPC showed a longer branch length than HCLV in the ML tree (Fig. 3). The difference may be due to the different passages in rabbits, with over 800 passages for LPC and over 400 for HCLV [15, 39]. Unfortunately, the complete sequence of the envelope glycoprotein gene of the Rovac strain is not available for analysis.

Thus far only a limited number of CSFV sequences representing the complete envelope glycoprotein gene have been deposited in GenBank. Given that currently available full-length CSFV genomic sequences are very diverse, previous attempts to amplify internal parts of the Paderborn strain genome using PCR primers that were designed based on the highly conserved regions of the available CSFV sequences have been unsuccessful. Consequently, primer walking has been used for PCR amplification and sequencing [22]. To detect a broad range of CSFV, we employed seven sets of PCR primers (Table 2) that were able to simultaneously amplify and sequence at least five distinct subgroups of CSFV based on the complete envelope glycoprotein gene. Using these primers was easy to amplify and sequence the region of complete envelope glycoprotein gene from distinct genotypes. It would be helpful to analyze the phylogenetic diversity from the complete envelope protein gene of CSFV around the world.

In this study we first reported on the use of E^{rns} and the complete envelope glycoprotein genes of CSFV for the phylogenetic analysis and suggested that the E2 gene should be expanded to the length at least 400 bp for a better bootstrap support. Our results showed that choosing appropriate genomic regions and fragment length for the phylogenetic analysis should provide results with improved discrimination. We also discovered a switch in genotypes of CSFV from the endemic to the introduced strains in Taiwan and postulated that the Taiwanese subgroup 2.1a viruses and the Paderborn strain possibly originated from somewhere in Asia.

Acknowledgements

The authors would like to thank all of those who submitted CSF tissue samples and Dr. Fan Lee (Animal Health Research Institute) for reviewing the manuscript. This work was supported by Council of Agriculture, Taiwan, Republic of China.

References

1. Bartak P, Greiser-Wilke I (2000) Genetic typing of classical swine fever virus isolates from the territory of the Czech Republic. Vet Microbiol 77: 59–70
2. Bjorklund HV, Stadejek T, Vilcek S, Belak S (1998) Molecular characterization of the 3′ noncoding region of classical swine fever virus vaccine strains. Virus Genes 16: 307–312
3. Bjorklund H, Lowings P, Stadejek T, Vilcek S, Greiser-Wilke I, Paton D, Belak S (1999) Phylogenetic comparison and molecular epidemiology of classical swine fever virus. Virus Genes 19: 189–195
4. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle

5. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791
6. Fritzemeier J, Teuffert J, Greiser-Wilke I, Staubach H, Moenning V (2000) Epidemiology of classical swine fever in Germany in the 1990s. Vet Microbiol 77: 29–41
7. Greiser-Wilke I, Depner K, Fritzemeier J, Haas L, Moennig V (1998) Application of a computer program for genetic typing of classical swine fever virus isolates from Germany. J Virol Methods 75: 141–150
8. Greiser-Wilke I, Fritzemeier J, Koenen F, Vanderhallen H, Rutili D, De Mia GM, Romero L, Sanchez-Vizcaino JM, San Gabriel A (2000) Molecular epidemiology of a large classical swine fever epizootic in the European Community in 1997–1998. Vet Microbiol 77: 17–27
9. Harasawa R, Giangaspero M (1999) Genetic variation in the 5′ end and NS5B regions of classical swine fever virus genome among Japanese isolates. Microbiol Immunol 43: 373–379
10. Hofmann MA, Brechtbuhl K, Stauber N (1994) Rapid characterization of new pestivirus strains by direct sequencing of PCR-amplified cDNA from the 5′ noncoding region. Arch Virol 139: 217–229
11. Ishikawa K, Nagai H, Katayama K, Tsuttsui M, Tanabayashi K, Takeuchi K, Hishiyama M, Saitoh A, Gotoh K, Muramtsu M, Yamada A (1995) Comparison of the entire nucleotide and deduced amino acid sequences of the attenuated hog cholera vaccine strain GPE$^-$ and the wild-type parental strain ALD. Arch Virol 140: 1385–1391
12. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120
13. Knowles NJ, Samuel AR (2003) Molecular epidemiology of foot-and-mouth disease virus. Virus Res 91: 65–80
14. Lee RCT (1954) Lapinized hog cholera vaccine in Taiwan. Scientific Agri (Taiwan) 2: 4–14 (In Chinese with English abstract)
15. Lin TTC, Lee RCT (1981) An overall report on the development of a highly safe and potent lapinized hog cholera virus strain for hog cholera control in Taiwan. National Science Council Special Publication Number 5, pp 1–42
16. Liu ST, Li RN, Wang DC, Chang SF, Chiang SC, Ho WC, Chang YS, Lai SS (1991) Rapid detection of hog cholera virus in tissues by the polymerase chain reaction. J Virol Methods 35: 227–236
17. Lowings JP, Paton DJ, Sands JJ, De Mia GM, Rutili D (1994) Classical swine fever: genetic detection and analysis of differences between virus isolates. J Gen Virol 75: 3461–3468
18. Lowings P, Ibata G, Needham J, Paton D (1996) Classical swine fever virus diversity and evolution. J Gen Virol 77: 1311–1321
19. Lowings P, Ibata G, De Mia GM, Rutili D, Paton D (1999) Classical swine fever in Sardinia: epidemiology of recent outbreaks. Epidemiol Infect 122: 553–559
20. Meyers G, Rumenapf T, Thiel HJ (1989) Molecular cloning and nucleotide sequence of the genome of classical swine fever virus. Virology 171: 555–567
21. Meyers G, Thiel HJ (1996) Molecular characterization of pestiviruses. Adv Virus Res 47: 53–118
22. Oleksiewicz MB, Rasmussen TB, Normann P, Uttenthal A (2003) Determination of sequence of the complete open reading frame and the 5′NTR of the Paderborn isolate of classical swine fever virus. Vet Microbiol 92: 311–325
23. Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357–358

24. Paton DJ, McGoldrick A, Greiser-Wilke I, Parchariyanon S, Song JY, Liou PP, Stadejek T, Lowings JP, Bjorklund H, Belak S (2000) Genetic typing of classical swine fever virus. Vet Microbiol 73: 137–157
25. Rumenapf T, Unger G, Strauss JH, Thiel HJ (1993) Processing of the envelope glycoproteins of pestivirus. J Virol 67: 3288–3294
26. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
27. Sakoda Y, Ozawa S, Damrongwatanapokin S, Sato M, Ishikawa K, Fukusho A (1999) Genetic heterogeneity of porcine and ruminant pestiviruses mainly isolated in Japan. Vet Microbiol 65: 75–86
28. Stadejek T, Warg J, Ridpath JF (1996) Comparative sequence analysis of the 5′ noncoding region of classical swine fever virus strains from Europe, Asia, and America. Arch Virol 141: 771–777
29. Stadejek T, Vilcek S, Lowings JP, Ballagi-Pordany A, Paton DJ, Belak S (1997) Genetic heterogeneity of classical swine fever virus in central Europe. Virus Res 52: 195–204
30. Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree toplogies. Mol Biol Evol 13: 964–969
31. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. PNAS 94: 6815–6819
32. Thiel HJ, Stark R, Weiland E, Rumenapf T, Meyers G (1991) Hog cholera virus: molecular composition of virions from a pestivirus. J Virol 65: 4705–4712
33. Tu C, Lu Z, Li H, Yu X, Liu X, Li Y, Zhang H, Yin Z (2001) Phylogenetic comparison of classical swine fever virus in China. Virus Res 81: 29–37
34. Vilcek S, Stadejek T, Ballagi-Pordany A, Lowings JP, Paton DJ, Belak S (1996) Genetic variability of classical swine fever virus. Virus Res 43: 137–147
35. Vilcek S, Paton DJ (1998) Application of genetic methods to study the relationship between classical swine fever outbreaks. Res Vet Sci 65: 89–90
36. Vilcek S, Paton D, Lowings P, Bjorklund H, Nettleton P, Belak S (1999) Genetic analysis of pestiviruses at the 3′ end of the genome. Virus Genes 18: 107–114
37. Widjojoatmodjo MN, van Gennip HP, de Smith AJ, Moormann RJ (1999) Comparative sequence analysis of classical swine fever virus isolates from the epizootic in The Netherlands in 1997–1998. Vet Microbiol 66: 291–299
38. Wong ML, Peng BY, Liu JJ, Chang TJ (2001) Cloning and sequencing of full-length cDNA of classical swine fever virus LPC strain. Virus Genes 23: 187–192
39. Wu HX, Wang JF, Zhang CY, Fu LZ, Pan ZS, Wang N, Zhang PW, Zhao WG (2001) Attenuated lapinized Chinese strain of classical swine fever virus: complete nucleotide sequence and character of 3′-noncoding region. Virus Genes 23: 69–76

Author's address: Prof. Shiow-Suey Lai, Department of Veterinary Medicine, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan; e-mail: lai@ccms.ntu.edu.tw

家畜衛試所研報No.40：90~110(2004)

台灣豬瘟病毒之親緣演化樹分析

潘居祥 1,2, 鍾明華 1, 黃天祥 1, 劉信孚 3, 林士鈺 1, 賴秀穗* 2
行政院農業委員會家畜衛生試驗所
國立台灣大學獸醫學研究所
馬偕醫院醫學研究部

摘要

將 1989 至 2003 年收集到的 158 個豬瘟病例針對 E^{rns} 及 E2 封套醣蛋白基因，以 RT-PCR 增幅此兩區間並進行核酸定序及親緣演化樹分析。雖然兩者出現類似的樹形圖，但 E^{rns} 比 E2 具有更好的區別效果。親緣演化樹分析結果顯示，115 個田間分離株歸屬於 2.1 亞群，又可進一步區分為 2.1a 及 2.1b，且此兩亞群都被認為是外來型病毒株，其餘的 43 個田間分離株屬於 3.4 亞群，被認為是本土型病毒株。2.1a 亞群病毒株最早發現於 1994 年，且於 1995 年以後成為田間優勢族群，然而 3.4 亞群病毒株盛行於 1994 年以前，但自 1996 年以後就無法從田間分離到。過去十餘年，我們戲劇性地看到台灣田間流行的豬瘟病毒從 3.4 亞群轉變成 2.1a 亞群，但不是從本土型病毒株的基因突變所造成。從基因庫資料分析發現，2.1a 亞群與德國分離株 Paderborn 及寮國分離株 L67 最相近，然而 2.1b 亞群與中國廣西省分離株最為相近。另針對完整的封套醣蛋白 E^{rns} 、E1 及 E2 以樹形圖分析 27 個豬瘟病毒核酸序列，結果發現此區間相較於全長的開放讀碼區(ORF)更具有良好的區別效果。

關鍵詞：豬瘟病毒、瘟疫病毒、親緣演化樹分析、基因分型

*抽印本索取作者
本文轉載自 Archives of Virology