ひ　ろ　ば

が麻疹発症した。同一車両内で感染した可能性が考えられた[15]。

麻疹全数把握制度による2006年の報告数は59例（疑い例を含む）で、そのうち定点からの報告が31例、定点以外からの報告が28例。このうち、病原体検査や血清学的検査により、麻疹が確定した症例が18例、否定された症例が40例、判定不可が1例であった。確定症例18例の年齢は、0～4歳　5例、5～9歳　2例、15～19歳　7例、20～29歳　3例。ワクチン接種歴は、未接種13例、接種済み4例、不明1例。

以上の報告で、今回の麻疹流行は、東京や埼玉からの移入感染例に端を発するものであったとし、今後も移入や輸入症例が流行の発端となる可能性が考えられ注意が必要であり、麻疹全数把握制度は有効な方法であり、継続する必要があることを強調している。

北海道の麻疹ゼロプロジェクト

小児科定点からの麻疹報告数は、表のごとく順調に減少しており、特に札幌市では平成17年に0となり、平成18年も0が続いていたが、11月30日～12月1日まで東京に出張していた31歳の男性が12月9日に麻疹発症、その後同僚や医療機関で感染し更に家族内感染もあり、平成18年12月に4例、平成19年1月に5例、合計9例の発症があった。成人5例、小児4例（9[illegible]、1歳、9ヵ月）。1例を除きワクチン未接[illegible]った[17]。

表

	平成13年	平成14年	平成15年	平成16年	平成17年
北海道	3,263	294	215	44	5
札幌市	925	22	118	1	0

文　献

1）世界の麻疹死亡者数が6割減：日経メディカルオンライン　2007.1.22.
2）井手：東京都医師会雑誌，Vol.50. No.7. 2003.
3）IASR：Vol.24. No.1. 2003.
4）IASR：Vol.24. No.8.
5）IASR：Vol.24. No.10.
6）IASR：Vol.25. No.2. 2004.
7）IASR：Vol.25. No.3.
8）IASR：＜特集＞麻疹　2001～2003：Vol.25. No.3. 2004.
9）IASR：＜特集＞麻疹・風疹　2006年現在；Vol.27. No.4. 2006.
10）関東における麻疹の集団発生：IDWR：Vol.18. No.16.
11）麻疹IgG抗体のavidity測定の臨床的意義：IASR；Vol.25. No.3.
12）[illegible]、阿部（千葉県こども病院）IASR：2006 8. 1.
13）高校における麻疹集団感染[illegible]、IASR：Vol.27. No.9.
14）東京都福祉保健局、東京都教育庁、その他各学校のホームページ.
15）沖縄の麻疹流行状況：IASR：Vol.28. No.5. 2007.
16）北海道ゼロプロジェクト：IASR：Vol.25. No.3. 2004.
17）富樫武弘：厚生労働省ワクチン研究班総会資料. 平成19年3月.

Journal of General Virology (2007), 88, 932–941 DOI 10.1099/vir.0.82495-0

Genomic analysis of diverse rubella virus genotypes

Yumei Zhou,[1,2] Hiroshi Ushijima[2] and Teryl K. Frey[1]

Correspondence
Teryl K. Frey
tfrey@gsu.edu

[1]Department of Biology, Georgia State University, Atlanta, GA, USA

[2]Department of Developmental Medical Sciences, Institute of International Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Based on the sequence of the E1 glycoprotein gene, two clades and ten genotypes of *Rubella virus* have been distinguished; however, genomic sequences have been determined for viruses in only two of these genotypes. In this report, genomic sequences for viruses in an additional six genotypes were determined. The genome was found to be well conserved. The viruses in all eight of these genotypes had the same number of nucleotides in each of the two open reading frames (ORFs) and the untranslated regions (UTRs) at the 5′ and 3′ ends of the genome. Only the UTR between the ORFs (the junction region) exhibited differences in length. Of the nucleotides in the genome, 78% were invariant. The greatest observed distance between viruses in different genotypes was 8.74% and the maximum calculated genetic distance was 14.78 substitutions in 100 sites. This degree of variability was similar among regions of the genome with two exceptions, both within the P150 non-structural protein gene: the N-terminal region that encodes the methyl/guanylyltransferase domain was less variable, whereas the hypervariable domain in the middle of the gene was more divergent. Comparative phylogenetic analysis of different regions of the genome was done, using sequences from 43 viruses of the non-structural protease (near the 5′ end of the genome), the junction region (the middle) and the E1 gene (the 3′ end). Phylogenetic segregation of sequences from these three genomic regions was similar with the exception of genotype 1B viruses, among which a recombinational event near the junction region was identified.

Received 24 August 2006
Accepted 20 November 2006

INTRODUCTION

Rubella virus is an important human pathogen that causes an acute, contagious disease known as rubella, 3-day measles or German measles, and severe birth defects (known as congenital rubella syndrome) when infection occurs during the first trimester of pregnancy (Chantler *et al.*, 2001). *Rubella virus* is the single member of the genus *Rubivirus* in the family *Togaviridae* and is an enveloped, single-stranded, positive-polarity RNA virus with a genome of approximately 10 kb. The genome contains two long open reading frames (ORFs): the 5′-proximal ORF (NSP-ORF) encodes two non-structural proteins, P150 and P90, that function in RNA replication, and the 3′-proximal ORF (SP-ORF) encodes three structural proteins: the capsid protein, C, and two envelope glycoproteins, E1 and E2. The SP-ORF is translated from a subgenomic RNA synthesized in infected cells (Frey, 1994). The genome also contains untranslated regions (UTRs) at its 5′ and 3′ ends and between the ORFs (known as the junction region).

Although rubella occurs worldwide, vaccination efforts with live-attenuated vaccines have been concentrated in developed countries. Currently, approximately 50% of countries have national vaccination efforts against rubella (Robertson *et al.*, 2003). Isolation and genetic sequencing of rubella viruses has been most thorough in countries pursuing elimination (Bosma *et al.*, 1996; Frey *et al.*, 1998; Icenogle *et al.*, 2006; Katow, 2004; Katow *et al.*, 1997a, b; Reef *et al.*, 2002; Saitoh *et al.*, 2006); however, collections have recently been assembled from other regions of the world (Donadio *et al.*, 2003; Katow, 2004; Zheng *et al.*, 2003a, c). Recently, a standard taxonomy for rubella viruses was adopted based on sequences of a standard window within the E1 gene and supported by sequencing of the SP-ORF of selected viruses (WHO, 2005). The taxonomy consists of two clades [corresponding to the previous genotypes I and II (Frey *et al.*, 1998; Zheng *et al.*, 2003a)] containing a total of ten genotypes, seven in clade 1 (1a, 1B, 1C, 1D, 1E, 1F and 1g) and three in clade 2 (2A, 2B and 2c); the genotypes designated in lower case are provisional. Within the E1 gene, maximal variation among clade 1 viruses is 5.8%, that among clade 2 viruses is 8.0%, and it is 8.2% between the two clades (Zheng *et al.*, 2003a). Geographically, clade 1 viruses circulate worldwide, whilst clade 2 viruses thus far have been restricted to Eurasia.

Thus far, ten complete genomic sequences of *Rubella virus* have been reported, which represent only two of the ten

Supplementary tables are available in JGV Online.

genotypes (eight sequences of genotype 1a viruses and two sequences of genotype 2A viruses). Among these sequences, genomic genetic variability is similar to that in the E1 gene, with the exception of a 'hypervariable region' (HVR) of greater variability in the middle of the P150 non-structural protein gene (Hofmann *et al.*, 2003; Zheng *et al.*, 2003b). Given the lack of representation of the majority of the genotypes in the current genomic database, the first goal of this study was to expand the number of genomic sequences, using viruses in our collection from six additional genotypes (1B, 1C, 1D, 1E, 2B and 2C). The second goal of this study was to extend phylogenetic analysis to 5′ regions of the genome, which had not previously been done. To this end, the sequence of the non-structural protease-encoding region within the P150 gene was determined and compared phylogenetically with the sequences of the junction region and the E1 gene from 43 viruses representing eight genotypes.

METHODS

Viruses, cells, RNA extraction, cDNA amplification and DNA sequencing. The viruses analysed in this study are listed in Table 1 (genomic sequences) and Supplementary Table S1, available in JGV Online (genomic regions). The Cba strain was provided by Dr Marta Zapata, University of Cordoba, Cordoba, Argentina. A monolayer of Vero cells (25 cm^2 T-flask or 60 mm plate) was infected with each virus. Three to five days post-infection, the culture medium was removed and total cellular RNA was extracted by using Tri-Reagent (Molecular Research Center) using the manufacturer's protocol. The RNA extracted from one 25 cm^2 T-flask or one 60 mm plate was resuspended in 50 µl double-distilled H_2O and stored at −80 °C until use. cDNA was synthesized in a 20 µl total reaction volume reaction containing 5 µl denatured (95 °C, 5 min) RNA template, 4 µl 5 × Reverse Superscript buffer (Invitrogen), 4 µl 2.5 mM dNTPs, 1 µl 0.1 M dithiothreitol, 1 µl 4 µM 3′E1/808 reverse primer (5′-TTTTTTTTTCTATACAGCAAC-3′; T_9 followed by the complement of nt 9751–9762 of the rubella virus genome), 1 µl (40 units) RNasin (Promega) and 1 µl (200 units) Superscript reverse transcriptase III (Invitrogen). The reaction was incubated at 55 °C for 60 min and then stored at −20 °C prior to use in PCR. Each 50 µl PCR contained 25 µl 2 × GC buffer I&II (TaKaRa), 8 µl 2.5 mM dNTPs, 3 µl cDNA template, 1 µl 40 µM appropriate forward and reverse primers and 0.5 µl (2.5 units) LA *Taq* polymerase (TaKaRa). Cycling parameters were determined according to the manufacturer's (TaKaRa) recommendations. For genomic sequencing, 10–11 overlapping fragments encompassing the entire genome were amplified by using appropriate primer pairs. The primers used to amplify the genomic regions are listed in Supplementary Table S2, available in JGV Online. Amplified fragments were purified following agarose-gel electrophoresis by using a QIAquick gel extraction kit (Qiagen). Sequencing reactions were performed bidirectionally by using appropriate primers and cycle-sequencing kits (ABI PRISM BigDye Terminator v. 3.1; PE Applied Biosystems) and resolved by using a 3100 Genetic Analyzer (Applied Biosystems). The 5′ and 3′ ends of the genome were determined by using a 5′/3′ FirstChoice RLM-RACE kit (Ambion Inc.).

Sequence analysis. For cataloguing and storage, sequences were input into free online sequence-alignment software (ALIGN Query, GENESTREAM SEARCH network server IGH, Montpellier, France; http://xylian.igh.cnrs.fr/bin/align-guess.cgi). The assembled nucleotide sequences were aligned by using the CLUSTAL_W multiple sequence-alignment program version 1.8 (Henikoff & Henikoff, 1994) and the PileUp program in the GCG software package (Genetics Computer Group, version 11.0; Accelrys Inc.). The TN93 substitution model (Tamura & Nei, 1993) with discrete gamma-distributed rate heterogeneity with eight gamma rate categories (Yang, 1994) (TN93 + Γ model) was used as a substitution model for phylogenetic reconstruction, as it was found statistically to be the best fit for our datasets. Maximum-likelihood (ML) phylogenetic analysis

Table 1. Rubella virus genomic sequences used in this study

Virus	Isolation site and year	Genotype	GenBank accession no.	Reference
Fth_USA64	Connecticut, USA, 1964	1a	M15240	Dominguez *et al.* (1990)
RA27_USA64	Pennsylvania, USA, 1964	1a	L78917	Pugachev *et al.* (1997)
M33_USA61	New Jersey, USA, 1961	1a	X05259, X72393	Clarke *et al.* (1987)
CEN_BEL63	Belgium, 1963	1a	AF188704	Lund & Chantler (2000)
TO-W_JAP67	Toyama, Japan, 1967	1a	AB047330	Kakizawa *et al.* (2001)
TO-V_JAP67	Toyama, Japan, 1967	1a	AB047329	Kakizawa *et al.* (2001)
ULR_GER84	Leipzig, Germany, 1984	1a	AF435865	Hofmann *et al.* (2003)
SUR_SVK74	Bratislava, Slovakia, 1974	1a	AF435866	Hofmann *et al.* (2003)
Cba_ARG88	Cordoba, Argentina, 1988	1B	DQ085339	This report
GUZ_GER92	Stuttgart, Germany, 1992	1B	DQ388280	This report
Anim_MEX97	Baja California, Mexico, 1997	1C	DQ085341	This report
JC2_NZL91	Auckland, New Zealand, 1991	1D	DQ388281	This report
6423_ITA97	Pavia, Italy, 1997	1E	DQ085343	This report
BRDII_CN80	Beijing, China, 1980	2A	AY258323	Zheng *et al.* (2003b)
BRI_CN79	Beijing, China, 1979	2A	AY258322	Zheng *et al.* (2003b)
AN5_KOR96	Seoul, Korea, 1996	2B	DQ085342	This report
I-11_ISR68	Tel-Aviv, Israel, 1968	2B	DQ085338	This report
C74_RUS97	Moscow, Russia, 1997	2c	DQ085340	This report
C4_RUS67	Moscow, Russia, 1967	2c	DQ388279	This report

was performed by using the TREE-PUZZLE program version 5.2 (Strimmer & von Haeseler, 1996). ML genetic distances and nucleotide-substitution statistical parameters were estimated under the selected TN93 + Γ substitution model with an initial neighbour-joining tree and then the best ML tree was reconstructed with these optimized parameters by using the quartet-puzzling method in the TREE-PUZZLE program. Fifty thousand and one hundred thousand quartet-puzzling steps were performed in constructing trees from the 19 genomic and 43 genomic region sequences, respectively. Sequence similarities and observed distances were calculated by using the Old Distance program in the GCG software package. A nucleotide sequence PLOTSIMILARITY plot across the genome (100 nt window) was generated by using the PLOTSIMILARITY program in the GCG software package. As the sequences of genotype 1a viruses were over-represented, the plot was generated by using six sequences from each clade, including members from each genotype. Nucleotide sequence substitution-rate analysis was carried out with PILEUP (GCG package), fastDNAml (version 1.2.2) and DNArates (version 1.1.0), employing default parameters. To detect recombination, phylogenetic analysis of sequences on either side of putative break points was conducted by using TREE-PUZZLE with the same parameter settings as were used in the genomic sequence analysis. Recombination was also analysed by using the sequence recombination-detection programs TOPALi (Milne *et al.*, 2004), RIP 2.0 (Recombination Identification Program; http://hivweb.lanl.gov/RIP/RIPsubmit.html) and the four-cluster likelihood mapping analysis in the TREE-PUZZLE program.

RESULTS

Genomic sequences and comparisons

A representative rubella virus phylogenetic tree based on the standard E1 gene window recommended by the WHO (nt 8291–9469) and containing the reference viruses for each genotype and the ten viruses for which the genomic sequence has been determined is shown in Fig. 1. As the genomic sequences were from genotype 1a and 2A viruses, the genomic sequences of nine representative viruses from six additional genotypes were determined (Table 1). Among these 19 viruses, with three exceptions, the genomes were 9762 nt in length and consisted (5′–3′) of a 40 nt 5′ UTR, a 6351 nt NSP-ORF, a 120 nt junction region, a 3192 nt SP-ORF and a 59 nt 3′ UTR. All three exceptions were in the junction region: the genome of one of the genotype 1B viruses (GUZ_GER92) was 9760 nt in length because it had a deletion of 2 nt at positions 6480–6481 (between the end of the NSP-ORF and the SG RNA start site) and the genomes of both genotype 2B viruses were 9761 nt in length because they had a deletion of 1 nt at position 6422 (between the SG RNA start site and the start of the SP-ORF). Fig. 2 shows a similarity plot across 12 genomic sequences proportionally representing all eight genotypes. Overall variability averaged approximately 7 % and was roughly comparable across the genome, with the exceptions of the 5′-terminal approximately 400 nt, a region encoding the methyl/guanylyltransferase (MT) domain within the P150 gene that exhibited variability of approximately 4 %, and a region of the P150 gene encompassing nt 2100–2400, the HVR, in which local variability peaked at up to 18 %. Pairwise observed genomic distances between viruses in different genotypes (see Supplementary Table S3, available in JGV Online) ranged from 2.0 to 8.7 %. The range of pairwise observed distances for genomic regions (the five genes and domains within P150 and P90) and the 3′ *cis*-acting element (3′CAE) is given in Table 2. Maximal observed distances of these regions were shown to range from 8.29 to 11.44 %, with the exception of the MT domain (5.24 %) and the HVR (21.18 %). Given this limited degree of variability across most of the genome, it was not suprising that 78 % of the nucleotides in the genome were invariant across the 19 sequences (Table 2).

The additional sequences contributed by this report greatly expand information on the genomic diversity of rubella viruses and, therefore, we took the opportunity to calculate a number of evolutionary parameters (Table 2). The maximum pairwise genetic distance (see Supplementary Table S3, available in JGV Online) was 14.78 substitutions in 100 sites, greater than the largest observed distance (8.74 observed substitutions in 100 sites). The maximal genetic distances for the genomic regions (the five genes, domains within P150 and P90 and the 3′CAE) ranged from 13.97 to 23.0 substitutions in 100 sites, with the exception of the MT domain (6.77 substitutions in 100 sites) and the HVR (35.85 substitutions in 100 sites). The transition/transversion site parameter (K) was 7.04 for the entire genome and ranged among genomic regions from 4.5 (HVR) to 13.35 (3′CAE). The pyrimidine and purine (Y/R) transition parameter of entire genome was 2.7 and varied among the genomic regions from 0.93 (X domain of the P150 gene) to 3.7 (C gene). To test whether the observed distance simply underestimates the genetic distance or whether substitution saturation has been reached, the pairwise number of transitions and transversions was plotted as a function of the calculated genetic distance (by using DAMBE; Xia & Xie, 2001), with the result that both transitions and transversions increased linearly with genetic distance, with the number of transitions being higher than transversions (data not shown). Neither reached a plateau, indicating that substitution saturation had not occurred.

Among the 19 genomic sequences, 78 % of the nucleotides were invariant. Not surprisingly, the parameter of rate heterogeneity, α, was 0.22 for the entire genome and varied between 0.19 and 0.33 across the genomic regions, with the exception of the HVR, within which $\alpha = 1.35$. These small α values indicated a strong substitution-rate heterogeneity among nucleotide sites across most of the genome (i.e. more than three-quarters of the nucleotides remained constant, whilst fewer than one-quarter exhibited variability). Within the HVR, 46 % of the nucleotides were variable.

Phylogenetic analysis

ML phylogenetic trees constructed from the complete genomic sequences, as well as from the NSP- and SP-ORFs, are displayed in Fig. 3. As in the E1-based tree, the six clade 2 sequences formed a clear, consistent branching pattern with high support values in all three trees, indicating that

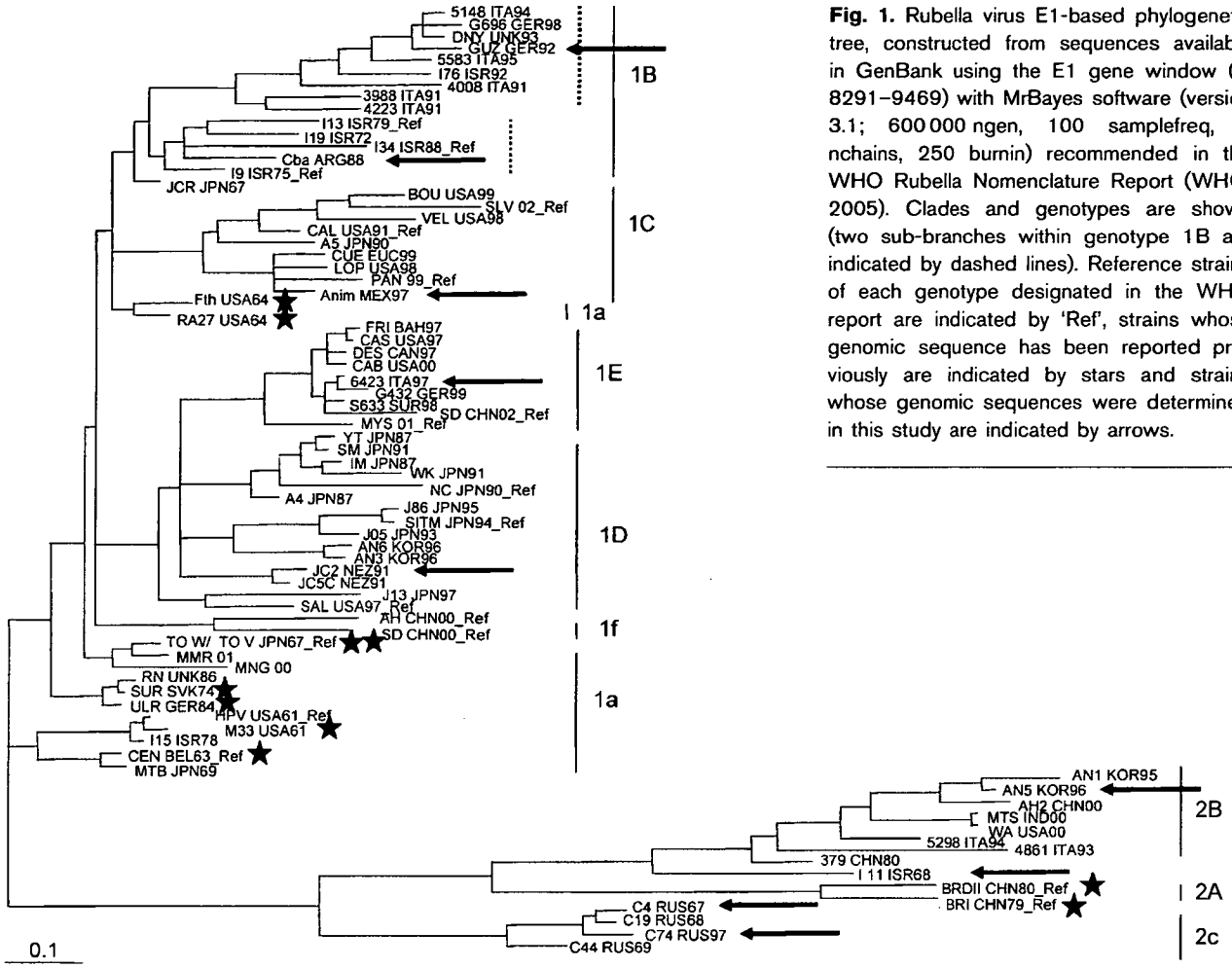

Fig. 1. Rubella virus E1-based phylogenetic tree, constructed from sequences available in GenBank using the E1 gene window (nt 8291–9469) with MrBayes software (version 3.1; 600 000 ngen, 100 samplefreq, 4 nchains, 250 burnin) recommended in the WHO Rubella Nomenclature Report (WHO, 2005). Clades and genotypes are shown (two sub-branches within genotype 1B are indicated by dashed lines). Reference strains of each genotype designated in the WHO report are indicated by 'Ref', strains whose genomic sequence has been reported previously are indicated by stars and strains whose genomic sequences were determined in this study are indicated by arrows.

genotypes 2A and 2B are related more closely to each other than to genotype 2c. In clade 1, the groupings of genotypes 1B, 1C, 1D and 1E on the three trees were consistent: the two genotype 1B sequences formed a branch, as did the individual genotype 1D and 1E sequences, whilst the individual genotype 1C sequence extended from the baseline with no relative relationship to other genotypes. On the genomic and SP-ORF trees, the eight genotype 1a sequences grouped into four pairs, indicated in Fig. 3 as a1 (TO-w and TO-v; a wild-type parent and the attenuated vaccine derived from it), a2 (Fth and RA27/3; both isolated in the north-eastern USA in 1964), a3 (CEN and M33; isolated from Europe and the USA in 1961–1962) and a4 (SUR and ULR; both isolated from eastern Europe in 1974 and 1984). Interestingly, on the NSP-ORF tree, M33 separated from CEN (a3) and clustered with the a4 grouping. We also constructed trees by using the sequences of the genes and regions within the NSP-ORF and SP-ORF (data not shown), with the result that they had the same general topology as the ORF-generated trees. The exception was the HVR-generated tree, on which each of the clade 1 viruses formed an individual branch, apart from the a1 and M33-SUR groupings, which were preserved.

Extensive phylogenetic analysis has not been done previously using sequences from 5′ regions of the genome. To do so, trees were constructed from the sequences of the non-structural protease (NP) region (nt 3035–3973; Fig. 2), the junction region and adjacent sequences (JR; nt 6351–6829, which includes the 3′ end of the P90 gene, the UTR between the ORFs and the 5′ end of the C gene) and the E1 gene [nt 8731–9469, the recommended window for routine genotyping (WHO, 2005)] of 43 viruses representing eight genotypes. ML phylogenetic trees constructed from these sequences are shown in Fig. 4(a). Clustering of viruses on the three trees was similar, with the exception of a group of seven genotype 1B viruses that formed a single branch on the NP tree and two branches (one of five and one of two viruses) on the JR tree, but did not form a cluster on the E1 tree. The single-nucleotide deletion at nt 6422 in the junction region detected in the genome sequence of the two genotype 2B viruses was confirmed in the JR sequences determined from three additional genotype 2B viruses, and the 2 nt deletion at nt 6480–6481 detected in the genome sequence of one of the two genotype 1B viruses was discovered in the JR sequences determined from four of the five additional genotype 1B viruses. Interestingly, this 2 nt

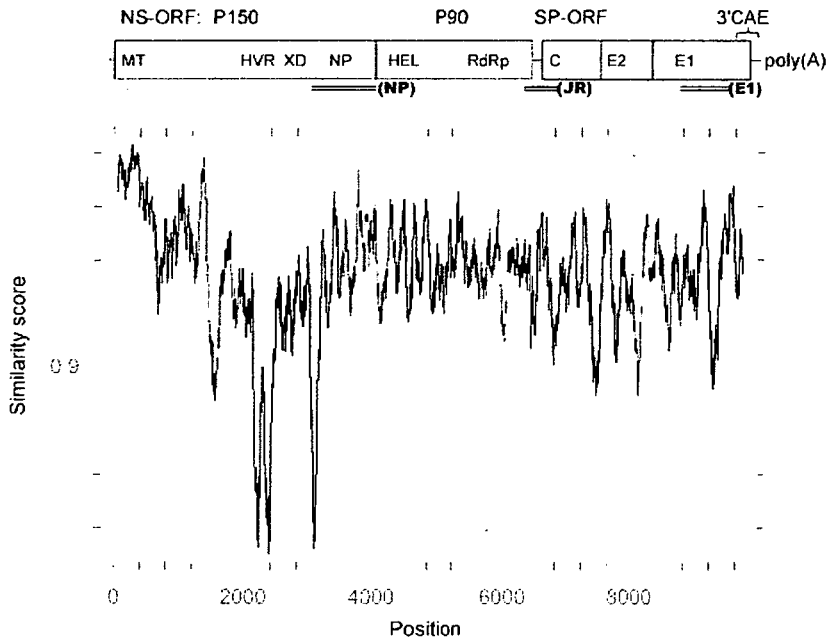

Fig. 2. Rubella virus genome nucleotide-similarity plot. The plot was generated from 12 genome sequences proportionally representing the two clades and eight genotypes by using the PLOTSIMILARITY program in the GCG software package with a 100 nt window. A genomic map is overlaid above the plot that shows the two ORFs, five genes, regions/domains within the NS-ORF (MT, methyl/guanylyltransferase; HVR, hypervariable region; XD, X domain; NP, non-structural protease; HEL, helicase; RdRp, RNA-dependent RNA polymerase or replicase) and the 3′ *cis*-acting elements (3′CAE). The regions from which sequences were determined from 43 viruses for comprehensive phylogenetic analysis (see Fig. 5) are also denoted (NP, non-structural protease; JR, junction region; E1, WHO standard E1 window).

Table 2. Maximum-likelihood parameters calculated for regions of the rubella virus genome

Region	G+C content (mol%)	No. sites	No. invariant sites (%)	Observed distance (range)*	Genetic distance (range)*	K*	Y/R*	α*
Genome†	69.5	9762	7611 (78.0)	0.22–8.74	0.22–14.78	7.04	2.7	0.22
P150†	71.8	3903	3033 (77.7)	0.26–9.3	0.26–15.89	5.97	2.25	0.23
P150: MT†	67.4	210	189 (90.0)	0–5.24	0–6.77	9.9	3.05	0.23
P150: HVR†	81.1	321	174 (54.2)	2.49–21.18	0.31–35.85	4.5	1.78	1.35
P150: XD†	74.2	507	377 (74.4)	0–11.44	0–23	7.2	0.93	0.24
P150: NP†	72.7	900	700 (77.8)	0.33–10	0.34–17.04	6.31	2.4	0.24
P90†	67.2	2448	1952 (79.7)	0.12–8.29	0.12–13.97	10.67	3.08	0.21
P90: HEL†	66.8	756	596 (78.8)	0.4–8.73	0.4–16.9	9.95	3.39	0.26
P90: RdRp†	66.9	1563	1250 (80.0)	0–8.83	0–15.3	11.06	2.74	0.19
C†	72.8	900	697 (77.4)	0.11–9.44	0.11–17.59	7.26	3.7	0.26
E2†	71	846	625 (73.9)	0–10.05	0–15.86	6	3.13	0.33
E1†	66.2	1446	1153 (79.7)	0.14–9.2	0.14–15.49	7.6	2.82	0.19
3′CAE†	64.9	307	244 (79.5)	0–8.79	0–14.67	13.35	2.98	0.2
NP‡	73	939	670 (71.4)	0–10.44	0–16.69	5.98	2.35	0.28
JR‡[c]	69.9	479	334 (69.7)	0–10.67	0–16.66	5.43	2.69	0.32
E1‡	67.5	739	546 (73.9)	0–10.42	0–21.06	7.74	3.42	0.21

*Parameters were calculated by using the TREE-PUZZLE program with setting of the TN93 substitution model with gamma discrete distribution-rate heterogeneity. K, Transitions/transversions; Y/R, pyrimidine transitions (T↔C)/purine transitions (A↔G); α, rate-heterogeneity shape parameter.

†Parameters were calculated from the indicated regions by using the 19 genomic sequences. P150: nt 41–3943; MT: methyltransferase, nt 230–439; HVR: hypervariable region, nt 2120–2440; XD: X domain, nt 2492–2998; NP: non-structural protease, nt 3041–3940; P90: nt 3944–6391; HEL: helicase region, nt 4043–4798; RdRp: RNA-dependent RNA polymerase region, nt 4826–6388; C, nt 6512–7411; E2: nt 7412–8257; E1: nt 8258–9703; 3′CAE: 3′ *cis*-acting elements, nt 9456–9762.

‡Parameters were calculated from the sequences of the NP (nt 3035–3973), JR (nt 6351–6829) and E1 (nt 8731–9469) regions determined for 43 viruses.

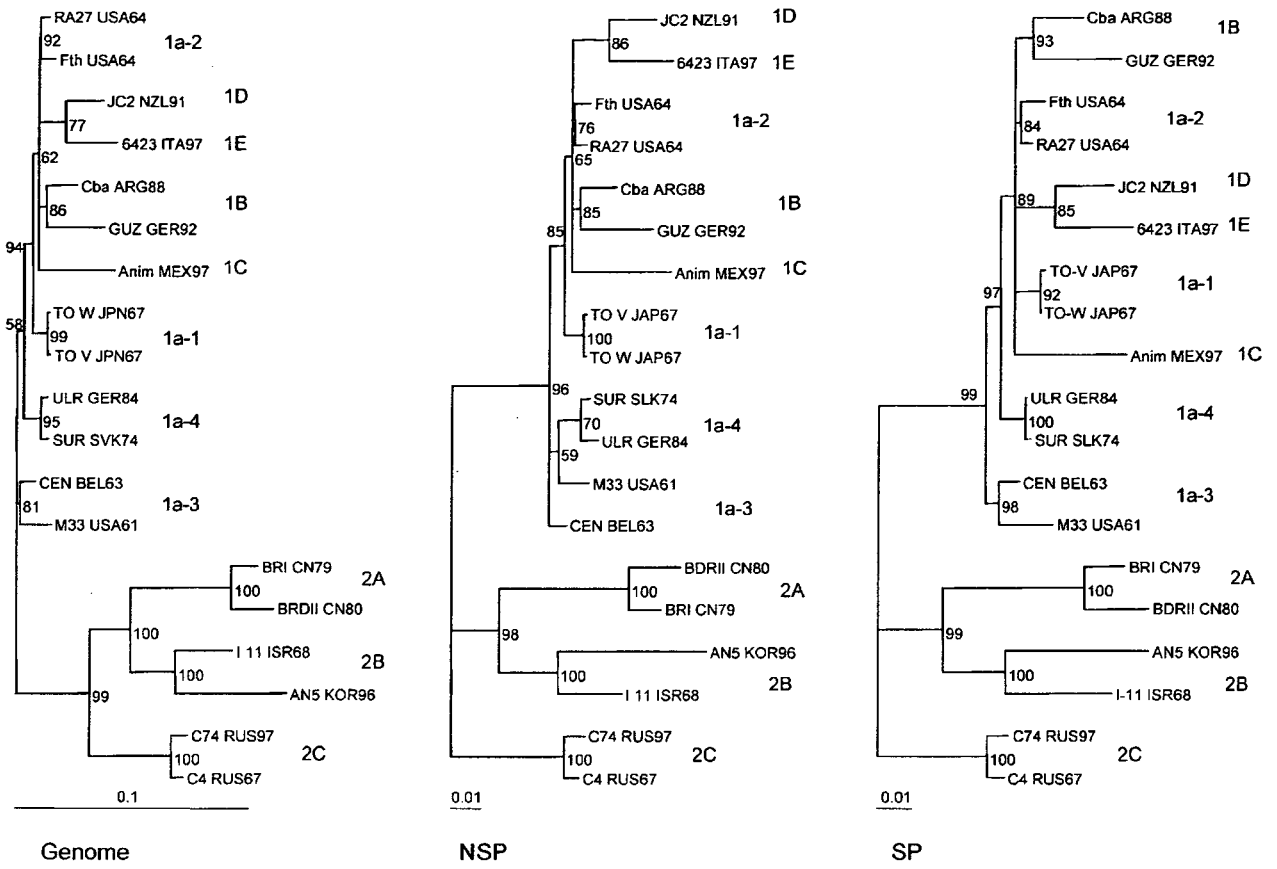

Fig. 3. Phylogenetic trees of genomic, NS-ORF and SP-ORF sequences. Trees were constructed by using TREE-PUZZLE (version 5.2) with 50 000 puzzle steps; reliability values are indicated on each node. Genotypes are denoted, including four subclusters of genotype 1a (1a-1 to 1a-4). Genetic distance (substitutions in 100 nt) calculated by using the TN93+Γ substitution model is indicated by the bar below each tree.

deletion did not co-segregate with the JR sequence of these viruses, as it was present in three members of the five-virus genotype 1B branch and both members of the two-virus genotype 1B branch [marked by an asterisk in Fig. 4(a)]. Evolutionary parameters calculated from these larger and more genotypically representative sequence sets, shown in Table 2, were similar to those calculated by using the smaller genome sequence set.

Detection of genomic recombination among genotype 1B viruses

The lack of co-segregation of the genotype 1B JR sequences and the 2 nt deletion led us to hypothesize that a recombinational event had occurred in this region of the genome of these viruses, at or downstream of the deletion. To test this hypothesis, we expanded the sequence determined upstream into the 3′ end of the P90 gene (the RdRp domain) and employed several software programs designed to detect recombination events, as well as secondary phylogenetic analysis, to detect putative break points. The RIP program predicted a break point at nt 6555, within the 5′ end of the C gene (which begins at nt 6512). As shown in Fig. 4(b), this prediction is supported by trees of sequences up- and downstream of this break point. Similar to the NP tree in Fig. 4(a), all seven genotype 1B sequences form a branch on the tree constructed by using sequences upstream of this break point (nt 5720–6554). On the tree constructed by using sequences downstream of the break point (nt 6555–6814), two sequences (TOM_UNK86 and 0754_GER92) were on a branch distinct from the other genotype 1B sequences, similar to the JR tree in Fig. 4(a). We thus conclude that a recombinational event occurred at or near this site during the evolution of these viruses.

DISCUSSION

The goal of this study was to expand the rubella virus genomic sequence database to include viruses in the majority of the currently defined genotypes. Whilst ten genomic sequences had been reported previously, only two of the ten currently defined genotypes were represented. This study added nine genomic sequences representing an additional six genotypes, encompassing the most widely

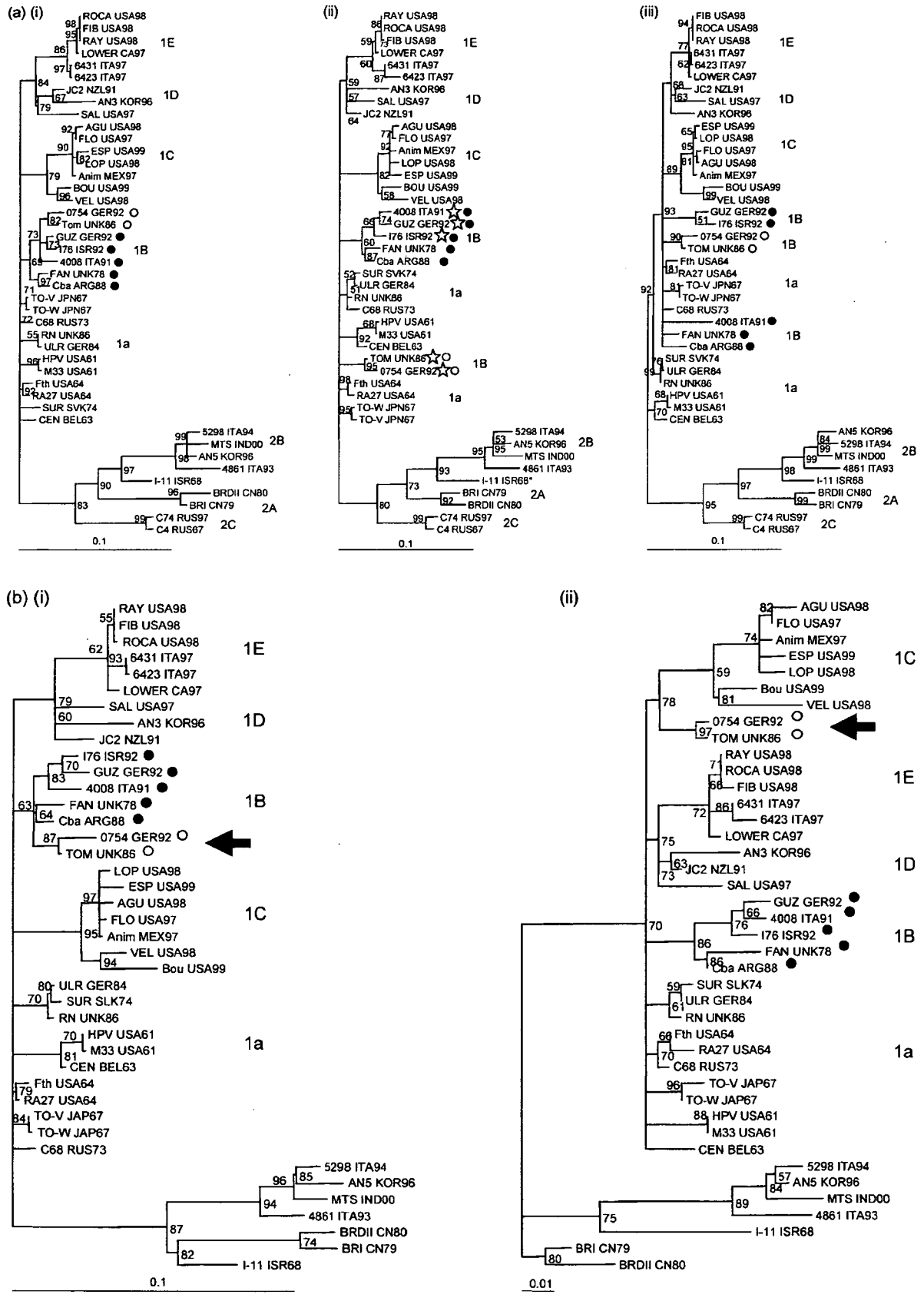
(a) (i)
ROCA USA98
FIB USA98
RAY USA98
LOWER CA97
6431 ITA97
6423 ITA97
JC2 NZL91
AN3 KOR96
SAL USA97
AGU USA98
FLO USA97
ESP USA99
LOP USA98
Anim MEX97
BOU USA99
VEL USA98
0754 GER92
Tom UNK86
GUZ GER92
I76 ISR92
4008 ITA91
FAN UNK78
Cba ARG88
TO-V JPN67
TO-W JPN67
C68 RUS73
RN UNK86
ULR GER84
HPV USA61
M33 USA61
Fth USA64
RA27 USA64
SUR SVK74
CEN BEL63
5298 ITA94
MTS IND00
AN5 KOR96
4861 ITA93
I-11 ISR68
BRDII CN80
BRI CN79
C74 RUS97
C4 RUS67
1E
1D
1C
1B
1a
2B
2A
2C
0.1
(ii)
RAY USA98
ROCA USA98
FIB USA98
LOWER CA97
6431 ITA97
6423 ITA97
AN3 KOR96
SAL USA97
JC2 NZL91
AGU USA98
FLO USA97
Anim MEX97
LOP USA98
ESP USA99
BOU USA99
VEL USA98
4008 ITA91
GUZ GER92
I76 ISR92
FAN UNK78
Cba ARG88
SUR SVK74
ULR GER84
RN UNK86
C68 RUS73
HPV USA61
M33 USA61
CEN BEL63
TOM UNK86
0754 GER92
Fth USA64
RA27 USA64
TO-W JPN67
TO-V JPN67
5298 ITA94
AN5 KOR96
MTS IND00
4861 ITA93
I-11 ISR68*
BRI CN79
BRDII CN80
C74 RUS97
C4 RUS67
1E
1D
1C
1B
1a
2B
2A
2C
0.1
(iii)
FIB USA98
ROCA USA98
RAY USA98
6431 ITA97
6423 ITA97
LOWER CA97
JC2 NZL91
SAL USA97
AN3 KOR96
ESP USA99
LOP USA98
FLO USA97
AGU USA98
Anim MEX97
BOU USA99
VEL USA98
GUZ GER92
I76 ISR92
0754 GER92
TOM UNK86
Fth USA64
RA27 USA64
TO-V JPN67
TO-W JPN67
C68 RUS73
4008 ITA91
FAN UNK78
Cba ARG88
SUR SVK74
ULR GER84
RN UNK86
HPV USA61
M33 USA61
CEN BEL63
AN5 KOR96
5298 ITA94
MTS IND00
4861 ITA93
I-11 ISR68
BRDII CN80
BRI CN79
C74 RUS97
C4 RUS67
1E
1D
1C
1B
1a
2B
2A
2C
0.1
(b) (i)
RAY USA98
FIB USA98
ROCA USA98
6431 ITA97
6423 ITA97
LOWER CA97
SAL USA97
AN3 KOR96
JC2 NZL91
I76 ISR92
GUZ GER92
4008 ITA91
FAN UNK78
Cba ARG88
0754 GER92
TOM UNK86
LOP USA98
ESP USA99
AGU USA98
FLO USA97
Anim MEX97
VEL USA98
Bou USA99
ULR GER84
SUR SLK74
RN UNK86
HPV USA61
M33 USA61
CEN BEL63
Fth USA64
RA27 USA64
TO-V JAP67
TO-W JAP67
C68 RUS73
5298 ITA94
AN5 KOR96
MTS IND00
4861 ITA93
BRDII CN80
BRI CN79
I-11 ISR68
1E
1D
1B
1C
1a
0.1
(ii)
AGU USA98
FLO USA97
Anim MEX97
ESP USA99
LOP USA98
Bou USA99
VEL USA98
0754 GER92
TOM UNK86
RAY USA98
ROCA USA98
FIB USA98
6431 ITA97
6423 ITA97
LOWER CA97
AN3 KOR96
JC2 NZL91
SAL USA97
GUZ GER92
4008 ITA91
I76 ISR92
FAN UNK78
Cba ARG88
SUR SLK74
ULR GER84
RN UNK86
Fth USA64
RA27 USA64
C68 RUS73
TO-V JAP67
TO-W JAP67
HPV USA61
M33 USA61
CEN BEL63
5298 ITA94
AN5 KOR96
MTS IND00
4861 ITA93
I-11 ISR68
BRI CN79
BRDII CN80
1C
1E
1D
1B
1a
0.01

Fig. 4. Phylogenetic trees based on sequence of genomic regions. (a) Trees were constructed using the sequence of the (i) non-structural protease (NP, nt 3035–3973), (ii) junction region (JR, nt 6351–6829, including the 3′ end of the NS-ORF, the UTR between the ORFs and the 5′ end of the C gene) and (iii) WHO standard E1 window (nt 8731–9469). On all three trees, genotype 1B viruses are indicated by dots (filled or empty for the clusters of five or two viruses, respectively, on the JR-based tree), and in (ii), genotype 1B viruses with a 2 nt deletion at nt 6480–6481 are indicated by stars. (b) Trees were constructed by using nt 5720–6554 (i) or 6555–6814 (ii), on either side of the putative recombination break point. Differential segregation of two of the genotype 1B viruses is indicated by arrows. All trees were constructed with TREE-PUZZLE (version 5.2; 100 000 puzzle steps). Reliability values are indicated on each node and genetic distance (substitutions in 100 nt) calculated by using the TN93+Γ substitution model is indicated by the bar below each tree.

divergent genotypes. The most striking finding was the genomic uniformity of rubella viruses, as it was discovered that 78 % of the nucleotides in the genomes of the viruses from the eight genotypes were invariant and these viruses preserved identical genomic dimensions across the two ORFs and two of the three UTRs. Only in the junction region (the UTR between the ORFs) of genotype 2B viruses, which had a 1 nt deletion, and a subset of genotype 1B viruses, which had a 2 nt deletion, was any plasticity observed. Such strict uniformity of genomic topology is highly unusual among RNA viruses (Huang *et al.*, 2004; Kang *et al.*, 2004; Kinney *et al.*, 1998; Saleh *et al.*, 2003; Takahashi *et al.*, 2003; Tarbatt *et al.*, 1997; van Cuyck *et al.*, 2003; Yang *et al.*, 2004). Sequence diversity was also low; across the eight genotypes, the maximum observed distance was <9 % and the maximum calculated genetic distance was 14.8 substitutions in 100 sites. Regardless of this difference, substitution saturation had not occurred, indicating that, despite the limited sequence diversity among rubella viruses, sufficient phylogenetic signal was retained to support the groupings observed (Salemi & Vandamme, 2003; Xia, 2000; Xia & Xie, 2001).

A sequence-similarity profile revealed, with two exceptions, a comparable pattern across the genome with local windows of similarity and dissimilarity varying about a relatively uniform mean, indicating that most genomic regions, including both virion protein and replicase protein genes, were equally divergent. Both observed and genetic distances between these genomic regions were comparable. The two exceptions were both within the P150 gene, with the N-terminal MT domain exhibiting less variability and the internal HVR exhibiting greater variability. Although the MT domain was predicted to encode both methyl- and guanylyltransferase activities (Rozanov *et al.*, 1992; neither activity has been demonstrated experimentally), the fact that 90 % of the nucleotide residues within this region are conserved raises the possibility that this region serves as a CAE in addition to encoding protein sequence. Consistent with this possibility, the phenotype of a cell culture-potentiating mutation discovered at nt 164 of the RA27/3 genome was found to be due to the nucleotide itself rather than to the encoded amino acid (Pugachev *et al.*, 2000). Conservation of the MT domain sequence has also been observed in other alpha-like family viruses (Gouvea *et al.*, 1998). The HVR encodes a proline- and arginine-rich domain of P150 termed the 'proline hinge' (Koonin *et al.*, 1992), although this domain contains several adaptor motifs that could serve to facilitate the association of P150 with other proteins. If this domain serves as a structural hinge between functional domains within the P150 protein, this could explain the lower constraint on sequence conservation within the HVR in comparison with the rest of the genome. On the other hand, Hofmann *et al.* (2003) reported data suggesting that the HVR among clade 1 viruses was under positive selection at the amino acid level. It should be pointed out that 'hypervariable region' is a relative term, in that HVRs in the genomes of other viruses are often more variable than the rubella virus HVR. For example, in the hepatitis E virus HVR, variability is >50 % (Arankalle *et al.*, 1999; Gouvea *et al.*, 1998; Nishizawa *et al.*, 2003; van Cuyck *et al.*, 2003).

With the exception of the HVR, nucleotide residues or sites across the genome showed a strong heterogeneity in rate of divergence, as indicated by the low value of the rate-heterogeneity parameter α. Sequence collections with low α values exhibit an L-shaped distribution on a graph of number of sites versus rate of divergence, rather than the bell-shaped curve generated when α is 1 (or >1). The low α value reflects the fact that roughly 80 % of the residues in the rubella virus genome were invariant in this collection of genomic sequences. The percentage of invariant residues at first and second codon positions was 93 %, compared with 48 % at third codon positions (Y. Zhou, unpublished data), and thus maintenance of amino acid sequence is a substantial component of the conservation of nucleotide sequence. Among third codon positions, the G+C content was 81 mol%, compared with 63 mol% among first and second codon positions (Y. Zhou, unpublished data), and thus there was selection for G and C residues. This selection was also evident in the HVR, within which the G+C content was 81 mol%, compared with 70 mol% for the genome.

Among the nucleotide substitutions at the 20 % of genomic sites that exhibited variability, transitions were strikingly more abundant than transversions; across the genome, the transition to transversion ratio, K, was 7.0 and varied among genomic regions from 4.5 to 13.4. Thus, the rubella virus genome exhibited the transition over transversion preference that has been well documented in DNA genomes (Meyer *et al.*, 1999; Salemi & Vandamme, 2003). This preference has been attributed to the facts that transitions are more likely to lead to silent mutations in amino acid

sequence and that, during replication, it is more likely that a mutation to a nucleotide of equal size (transition) will occur than to a nucleotide of different size (transversion). In RNA genomes, the possibility of both G–C and G–U pairing would also favour transitions in the replication process. Interestingly, pyrimidine transitions were favoured over purine transitions by a ratio (Y/R) of 2.7 across the entire genome; Y/R varied from 0.9 to 3.7 in genomic regions. Within the HVR, the most variable region of the genome, both K and Y/R were lower than for the entire genome and most of the other genomic regions, indicating that the variability in this region was generated in part by relaxing of the genomic preference for pyrimidine transitions over transversions and purine transitions.

Phylogenetic analysis of rubella viruses has traditionally been done on the basis of E1 gene or subE1 gene sequences and a standard taxonomy was proposed recently, based on a window within the E1 gene, that was substantiated by using complete SP-ORF sequences (WHO, 2005). The second goal of this study was to extend phylogenetic analysis to the 5′ region of the genome and we found that generally comparable trees, in terms of both overall variability and phylogenetic clustering, were generated with sequence windows in the NSP-ORF. The exception was a group of seven genotype 1B viruses that formed a branch in a tree based on NP sequence, but formed two branches on the basis of JR sequence. Intriguingly, a deletion in the junction region of five of these seven viruses did not segregate with the two phylogenetic branches on the JR tree. Analysis revealed a recombinational event, putatively near the 5′ end of the C gene, that led to the generation of the two branches on the JR tree. There was one previous report of a natural recombination event in *Rubella virus* (in the E1 gene; Zheng *et al.*, 2003a), but the origin of the recombinant strain was in doubt because one of the parents was related closely to a commonly used laboratory strain. Thus, this was the first conclusive evidence of rubella virus recombination in nature.

Interestingly, the E1 sequences of the seven genotype 1B viruses did not cluster on the E1-based tree and, in comparison with the NP- and JR-based trees, this could be due to divergence or additional recombinational events. As can be seen in the tree in Fig. 1, genotype 1B consists of two sub-branches that would not necessarily appear to be related if fewer sequences were employed (e.g. the E1-based tree in Fig. 4). It is also to be noted that all of the WHO reference strains are on one of these sub-branches. Thus, for this genotype, phylogenetic analysis using sequences from the NSP-ORF region of the genome could be useful in assessing relatedness.

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Institutes of Health (AI21389). We thank Xianfeng Chen for software assistance, Duping Zheng, Suganthi Suppiah and Hui Zhao for preliminary sequence determinations and Ping Jiang for processing sequencing reactions and gels.

REFERENCES

Arankalle, V. A., Paranjape, S., Emerson, S. U., Purcell, R. H. & Walimbe, A. M. (1999). Phylogenetic analysis of hepatitis E virus isolates from India (1976–1993). *J Gen Virol* **80**, 1691–1700.

Bosma, T. J., Best, J. M., Corbett, K. M., Banatvala, J. E. & Starkey, W. G. (1996). Nucleotide sequence analysis of a major antigenic domain of the E1 glycoprotein of 22 rubella virus isolates. *J Gen Virol* 77, 2523–2530.

Chantler, J. K., Wolinsky, J. S. & Tingle, A. (2001). Rubella virus. In *Fields Virology*, 4th edn, pp. 963–990. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott Williams & Wilkins.

Clarke, D. M., Loo, T. W., Hui, I., Chong, P. & Gillam, S. (1987). Nucleotide sequence and in vitro expression of rubella virus 24S subgenomic messenger RNA encoding the structural proteins E1, E2 and C. *Nucleic Acids Res* **15**, 3041–3057.

Dominguez, G., Wang, C. Y. & Frey, T. K. (1990). Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution. *Virology* **177**, 225–238.

Donadio, F. F., Siqueira, M. M., Vyse, A., Jin, L. & Oliveira, S. A. (2003). The genomic analysis of rubella virus detected from outbreak and sporadic cases in Rio de Janeiro state, Brazil. *J Clin Virol* 27, 205–209.

Frey, T. K. (1994). Molecular biology of rubella virus. *Adv Virus Res* **44**, 69–160.

Frey, T. K., Abernathy, E. S., Bosma, T. J., Starkey, W. G., Corbett, K. M., Best, J. M., Katow, S. & Weaver, S. C. (1998). Molecular analysis of rubella virus epidemiology across three continents, North America, Europe, and Asia, 1961-1997. *J Infect Dis* 178, 642–650.

Gouvea, V., Snellings, N., Popek, M. J., Longer, C. F. & Innis, B. L. (1998). Hepatitis E virus: complete genome sequence and phylogenetic analysis of a Nepali isolate. *Virus Res* 57, 21–26.

Henikoff, S. & Henikoff, J. G. (1994). Position-based sequence weights. *J Mol Biol* 243, 574–578.

Hofmann, J., Renz, M., Meyer, S., von Haeseler, A. & Liebert, U. G. (2003). Phylogenetic analysis of rubella virus including new genotype I isolates. *Virus Res* 96, 123–128.

Huang, F. F., Sun, Z. F., Emerson, S. U., Purcell, R. H., Shivaprasad, H. L., Pierson, F. W., Toth, T. E. & Meng, X. J. (2004). Determination and analysis of the complete genomic sequence of avian hepatitis E virus (avian HEV) and attempts to infect rhesus monkeys with avian HEV. *J Gen Virol* 85, 1609–1618.

Icenogle, J. P., Frey, T. K., Abernathy, E., Reef, S. E., Schnurr, D. & Stewart, J. A. (2006). Genetic analysis of rubella viruses found in the United States between 1966 and 2004: evidence that indigenous rubella viruses have been eliminated. *Clin Infect Dis* **43** (Suppl. 3), S133–S140.

Kakizawa, J., Nitta, Y., Yamashita, T., Ushijima, H. & Katow, S. (2001). Mutations of rubella virus vaccine TO-336 strain occurred in the attenuation process of wild progenitor virus. *Vaccine* 19, 2793–2802.

Kang, S. Y., Yun, S. I., Park, H. S., Park, C. K., Choi, H. S. & Lee, Y. M. (2004). Molecular characterization of PL97-1, the first Korean isolate of the porcine reproductive and respiratory syndrome virus. *Virus Res* 104, 165–179.

Katow, S. (2004). Molecular epidemiology of rubella virus in Asia: utility for reduction in the burden of diseases due to congenital rubella syndrome. *Pediatr Int* 46, 207–213.

Katow, S., Minahara, H., Fukushima, M. & Yamaguchi, Y. (1997a). Molecular epidemiology of rubella by nucleotide sequences of the rubella virus E1 gene in three East Asian countries. *J Infect Dis* **176**, 602–616.

Katow, S., Minahara, H., Ota, T. & Fukushima, M. (1997b). Identification of strain-specific nucleotide sequences in E1 and NS4 genes of rubella virus vaccine strains in Japan. *Vaccine* **15**, 1579–1585.

Kinney, R. M., Pfeffer, M., Tsuchiya, K. R., Chang, G. J. & Roehrig, J. T. (1998). Nucleotide sequences of the 26S mRNAs of the viruses defining the Venezuelan equine encephalitis antigenic complex. *Am J Trop Med Hyg* **59**, 952–964.

Koonin, E. V., Gorbalenya, A. E., Purdy, M. A., Rozanov, M. N., Reyes, G. R. & Bradley, D. W. (1992). Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci U S A* **89**, 8259–8263.

Lund, K. D. & Chantler, J. K. (2000). Mapping of genetic determinants of rubella virus associated with growth in joint tissue. *J Virol* **74**, 796–804.

Meyer, S., Weiss, G. & von Haeseler, A. (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103–1110.

Milne, I., Wright, F., Rowe, G., Marshall, D. F., Husmeier, D. & McGuire, G. (2004). TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**, 1806–1807.

Nishizawa, T., Takahashi, M., Mizuo, H., Miyajima, H., Gotanda, Y. & Okamoto, H. (2003). Characterization of Japanese swine and human hepatitis E virus isolates of genotype IV with 99% identity over the entire genome. *J Gen Virol* **84**, 1245–1251.

Pugachev, K. V., Abernathy, E. S. & Frey, T. K. (1997). Genomic sequence of the RA27/3 vaccine strain of rubella virus. *Arch Virol* **142**, 1165–1180.

Pugachev, K. V., Galinski, M. S. & Frey, T. K. (2000). Infectious cDNA clone of the RA27/3 vaccine strain of Rubella virus. *Virology* **273**, 189–197.

Reef, S. E., Frey, T. K., Theall, K., Abernathy, E., Burnett, C. L., Icenogle, J., McCauley, M. M. & Wharton, M. (2002). The changing epidemiology of rubella in the 1990s: on the verge of elimination and new challenges for control and prevention. *JAMA* **287**, 464–472.

Robertson, S. E., Featherstone, D. A., Gacic-Dobo, M. & Hersh, B. S. (2003). Rubella and congenital rubella syndrome: global update. *Rev Panam Salud Publica* **14**, 306–315.

Rozanov, M. N., Koonin, E. V. & Gorbalenya, A. E. (1992). Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J Gen Virol* **73**, 2129–2134.

Saitoh, M., Shinkawa, N., Shimada, S., Segawa, Y., Sadamasu, K., Hasegawa, M., Kato, M., Kozawa, K., Kuramoto, T. & other authors (2006). Phylogenetic analysis of envelope glycoprotein (E1) gene of rubella viruses prevalent in Japan in 2004. *Microbiol Immunol* **50**, 179–185.

Saleh, S. M., Poidinger, M., Mackenzie, J. S., Broom, A. K., Lindsay, M. D. & Hall, R. A. (2003). Complete genomic sequence of the Australian south-west genotype of Sindbis virus: comparisons with other Sindbis strains and identification of a unique deletion in the 3′-untranslated region. *Virus Genes* **26**, 317–327.

Salemi, M. & Vandamme, A.-M. (2003). *The Phylogenetic Handbook: a Practical Approach to DNA and Protein Phylogeny.* Cambridge: Cambridge University Press.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**, 964–969.

Takahashi, K., Kang, J. H., Ohnishi, S., Hino, K., Miyakawa, H., Miyakawa, Y., Maekubo, H. & Mishiro, S. (2003). Full-length sequences of six hepatitis E virus isolates of genotypes III and IV from patients with sporadic acute or fulminant hepatitis in Japan. *Intervirology* **46**, 308–318.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–526.

Tarbatt, C. J., Glasgow, G. M., Mooney, D. A., Sheahan, B. J. & Atkins, G. J. (1997). Sequence analysis of the avirulent, demyelinating A7 strain of Semliki Forest virus. *J Gen Virol* **78**, 1551–1557.

van Cuyck, H., Juge, F. & Roques, P. (2003). Phylogenetic analysis of the first complete hepatitis E virus (HEV) genome from Africa. *FEMS Immunol Med Microbiol* **39**, 133–139.

WHO (2005). Standardization of the nomenclature for genetic characteristics of wild-type rubella viruses. *Wkly Epidemiol Rec* **80**, 126–132.

Xia, X. (2000). *Data Analysis in Molecular Biology and Evolution.* Boston: Kluwer Academic Publishers.

Xia, X. & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* **92**, 371–373.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306–314.

Yang, D. K., Kim, B. H., Kweon, C. H., Kwon, J. H., Lim, S. I. & Han, H. R. (2004). Molecular characterization of full-length genome of Japanese encephalitis virus (KV1899) isolated from pigs in Korea. *J Vet Sci* **5**, 197–205.

Zheng, D. P., Frey, T. K., Icenogle, J., Katow, S., Abernathy, E. S., Song, K. J., Xu, W. B., Yarulin, V., Desjatskova, R. G. & other authors (2003a). Global distribution of rubella virus genotypes. *Emerg Infect Dis* **9**, 1523–1530.

Zheng, D. P., Zhou, Y. M., Zhao, K., Han, Y. R. & Frey, T. K. (2003b). Characterization of genotype II rubella virus strains. *Arch Virol* **148**, 1835–1850.

Zheng, D. P., Zhu, H., Revello, M. G., Gerna, G. & Frey, T. K. (2003c). Phylogenetic analysis of rubella virus isolated during a period of epidemic transmission in Italy, 1991-1997. *J Infect Dis* **187**, 1587–1597.

Progress in Medicine 第12号 別刷

Vol.27 No.12 2007.12

原　著

国産狂犬病ワクチンの皮内接種によるヒトへの狂犬病曝露前免疫の検討

柳澤　如樹　　髙山　直秀　　菅沼　明彦

ライフ・サイエンス

Prog.Med.
27：2937～2939, 2007

原　著

国産狂犬病ワクチンの皮内接種によるヒトへの狂犬病曝露前免疫の検討

Yanagisawa Naoki　Takayama Naohide　Suganuma Akihiko
柳澤　如樹[1)]　高山　直秀[2)]　菅沼　明彦[1)]

目　的

2006年11月に国内で相次いで輸入狂犬病患者が発生した後[1,2)]，狂犬病ワクチンの需要が急増した．しかし，需要の増加が供給を上回っていたため，全国的に狂犬病ワクチンが品不足に陥った．このため，厚生労働省は緊急避難的に狂犬病曝露前免疫を制限する方針を示した．世界保健機関(WHO)は，狂犬病流行地において動物による咬傷を受けた場合，抗狂犬病免疫グロブリン(RIG)の投与と組織培養不活化ワクチン接種による曝露後発症予防を勧告している[3)]．ただし，曝露前免疫を受けていればRIGの投与が不要になる．RIGは世界的に不足しており，入手が容易でないため，曝露後免疫の効果を確実にする上で，狂犬病曝露前免疫を行っておくことが重要である．わが国における狂犬病ワクチンの生産量は少なく，急な増産ができないことを鑑みれば，今回のようなワクチン不足の事態に備えて，接種ワクチン量が少なくとも，高い効果を上げることができる接種法を検討しておくことが必要である．

狂犬病常在地であるタイでは，１人当たりの狂犬病ワクチン接種量を減量するために，皮内接種法(タイ赤十字方式)が広く用いられ，曝露後発症予防に効果を上げているばかりか[4)]，曝露前免疫にも採用され成果を上げている．国産の狂犬病ワクチンを用いての皮内接種法については，高山らの皮内・皮下併用法以外に，その有効性や安全性がこれまで検討されておらず[5,6)]，皮内接種のみによる曝露前免疫は１例が報告されているにすぎない[7)]．

われわれは，国産ワクチンを用いた皮内接種法の可否を知るため，同意を得た健康成人に対して，国産狂犬病ワクチンを皮内接種して，狂犬病曝露前免疫の効果と安全性を調査した．

対象と方法

１．対　象

本調査の目的，調査項目，接種ワクチンと予測される副反応について文書，および口頭で説明をして，同意が得られた医療・獣医療関係者17例を対象とした．

２．接種ワクチン

化学及血清療法研究所(化血研)製組織培養不活化狂犬病ワクチンのロットRB02およびRB03を用いた．狂犬病ワクチンは溶解液１mLで溶解した後，その0.1 mLずつを左右前腕に皮内注射した．

３．局所および全身反応

全例について，皮内接種15分後の接種局所における膨疹，発赤を視診で確認し，瘙痒感の有無を質問した．さらに，次回接種時および採血時に前回注射による局所の腫脹，発赤，疼痛，瘙痒感の自覚症状の有無について問診した．

４．抗体検査

狂犬病ワクチンを２～４週間隔で２回皮内接種し，２回目接種２～３週間後に採血して，血中抗狂犬病抗体価を測定した．血中抗狂犬病抗体価は，化血研臨床検査センターに依頼して，Platelia® rabies kit(BIO RAD Laboratories)を用いてELISA法で測定した．

１）東京都立駒込病院感染症科　２）東京都立駒込病院小児科

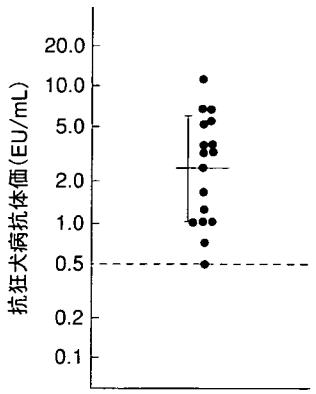

図1　被験者の抗狂犬病抗体価の分布

国産狂犬病ワクチン0.1mLを左右前腕に2～4週間隔で2回皮内接種した．2～3週間後の抗狂犬病抗体価(●)をELISA法で測定した．

横棒は幾何平均値，横棒を貫く縦棒は±SDを示す(n＝17)．

横点線はWHOの定める発症防御レベルを示す．

結　　果

1．対象者の年齢分布

対象者は男性11例，女性6例の合計17例であった．年齢分布は22歳から55歳まで，20歳代が13例，30歳代が2例，40歳代が1例，50歳代が1例であり，平均年齢は30.2±7.5歳であった．

2．血中抗狂犬病抗体価

ワクチンを2回皮内接種した2～3週間後には，17例全例で，抗体価が0.5 EU/mL以上に上昇していた．抗体価の範囲は0.5～11.5 EU/mL，幾何平均値は2.5 EU/mLであった(図1)．

3．接種後の局所反応および全身症状

ワクチン接種15分後，局所の発赤を呈した者は5例，腫脹を認めた者は9例，疼痛を認めた者は0例，瘙痒感を認めた者は2例であった．局所の発赤は，数日間残ったと報告した者があったが，発熱，頭痛，倦怠感などの全身症状を報告した例はなかった．

考　　察

本邦での狂犬病曝露前免疫は，組織培養不活化ワクチン1回量1.0 mLを4週間隔で2回，その後6～12カ月後に1回皮下注射する方式が標準である[8]．しかし，今回生じたようなワクチン供給不足の状況において，1人に1回量1.0 mLを投与すれば，短期間にワクチンが枯渇して，実際上狂犬病ワクチン接種が不可能となるであろう．高山らは，狂犬病ワクチン成分にアレルギー反応がある患者に皮内接種で曝露前免疫を行った例を報告している[7]．本報告では，患者に全身反応はなく，有意な抗体上昇がみられ，国産ワクチンを用いても皮内接種による曝露前免疫が可能であることが示唆された．

今回われわれは，ワクチン不足の状況下でも，できる限り多くの人々からの要望に応えられるように，狂犬病ワクチンの1回接種量を減量しても効果が得られる方法として，タイ赤十字方式を改変した皮内接種法(駒込方式)を考案し，その有効性と安全性を検討した．小規模な接種試験であったが，接種者全員の抗体価が0.5 EU/mL 以上となった．これはWHOの定める発症防御レベルの0.5 IU/mL以上と同等とみなすことができる．また，接種部位の発赤など軽度の局所副反応を認めたが，全身的副反応は認められなかった．

わが国における狂犬病ワクチンの生産量は，定期接種のワクチンに比べれば非常に少なく，急激な需要の増大が起これば，狂犬病ワクチンが不足する事態は避けられない．ワクチン不足への対処法として，ワクチンの備蓄や外国から緊急輸入などの方策が考えられる．しかし，ワクチンの備蓄は期限切れで廃棄することになる可能性や，外国産のワクチンの輸入に関しては法的な制限もあり，現実的には難しい．また，ワクチンは急に増産することができない．交通手段の発達により，日本から数時間以内で狂犬病常在地に渡航できることを考えると，今後も輸入狂犬病患者が発生し，再び狂犬病ワクチンの供給不足が発生する可能性があることを十分認識する必要がある．

今回検討した駒込方式は，接種量が少なくとも，標準法に劣らず効果がみられ，かつ副反応も軽微であるため，ワクチン不足時にはもちろん，平時にも使用できる有効な曝露前免疫法となる可能性があり，今後はより多くの被接種者を対象として，さらなる検討を行う価値がある．

本研究は厚生労働科学新興・再興感染症研究事業による研究費補助を受けた．

文　　献

1）山本舜悟，岩崎千尋，大野博司ほか：本邦36年ぶりの狂犬病輸入症例の報告―京都の事例．病原微生物検出情報　2007；28：63-64.

2）高橋華子，相楽裕子，藤田せつ子ほか：36年ぶりに国内で発症した狂犬病の臨床経過と感染予防策．病原微生物検出情報　2007；28：64-65.

3）Current WHO guide for rabies pre and post-exposure treatment in humans. Geneva：World Health Organization, 2002. http://www.who.int/rabies/en/WHO_guide_rabies_pre_post_exp_treat_humans.pdf
4）Chutivongse S, Wilde H, Supich C, et al：Postexposure prophylaxis for rabies with antiserum and intradermal vaccination. Lancet　1990；**335**：896-898.
5）高山直秀，万年和明，井戸田一朗ほか：ヒトへの皮内および皮下接種併用法による狂犬病曝露前免疫の検討．臨床とウイルス　2001；**29**：395-397.
6）高山直秀，井戸田一朗，加藤康幸：ヒトへの皮内・皮下接種併用法による狂犬病曝露前発病予防の検討．臨床とウイルス　2003：**31**：62-66.
7）高山直秀，大隅邦夫，作間　晋：狂犬病ワクチン成分に対する過敏反応のため皮内接種法により狂犬病曝露前免疫を行った1例．感染症学雑誌　1999；**73**：600-601.
8）木村三生夫，平山宗宏，堺　春美：予防接種の手引き第11版，近代出版，東京，2006；pp.327-331.

Pre-Exposure Prophylaxis for Rabies with Intradermal Injection Using Japanese Rabies Vaccine

Naoki Yanagisawa[1)], Naohide Takayama[2)] and Akihiko Suganuma[1)]

1）Department of Infectious Diseases, Tokyo Metropolitan Komagome Hospital
2）Department of Pediatrics, Tokyo Metropolitan Komagome Hospital

In November 2006, two consecutive imported rabies cases were reported in Japan. The demand for rabies vaccine has grown rapidly, resulting in a shortage of the vaccine. Therefore, pre-exposure prophylaxis for rabies was restricted. In order to prepare for the vaccine shortage, it is necessary to consider a method that is effective yet uses less amount of the vaccine. The intradermal method we tested uses only 20% of the vaccine dose required under the standard method, but every subject tested had a sufficient rise in their anti-rabies antibody titer. This was a small inoculation trial, but intradermal vaccination is an effective method, and may be used on a regular basis not only when vaccine is short.

改正結核予防法2005年施行後の全国BCGワクチン累積接種率調査

東京都立駒込病院小児科[1]，崎山小児科[2]，国立感染症研究所感染症情報センター[3]
髙山　直秀[1]　崎山　弘[2]　岡部　信彦[3]

日本小児科学会雑誌　第111巻　第8号別刷

日本小児科学会雑誌　111巻8号　1042～1044（2007年）

改正結核予防法2005年施行後の全国BCGワクチン累積接種率調査

東京都立駒込病院小児科[1]，崎山小児科[2]，国立感染症研究所感染症情報センター[3]
髙山　直秀[1]　崎山　弘[2]　岡部　信彦[3]

要　旨

2005年度よりBCGワクチンの接種対象年齢が，結核予防法改正前の「生後4歳に達するまで」から「生後6カ月に達するまで」に引き下げられた．法改正に伴いBCGワクチンの接種率が低下し，接種もれ者が増加することが懸念されたため，法改正後にBCGワクチン接種年齢に達した小児におけるBCGワクチンの全国累積接種率を調査した．改正法実施後の全国累積接種率は生後6カ月以前に約97%に達しており，約93%の接種対象者が生後3カ月から6カ月に達する前までに接種を受けていた．今後は接種漏れ者を早期に発見して接種を勧奨できる体制を整備するとともに，BCGワクチンの早期接種がジフテリア・百日咳・破傷風3種混合（DPT）ワクチン及びポリオ生ワクチンの接種率にどのような影響を与えているかを検証する必要がある．

キーワード：BCG，結核予防法，累積接種率，完了率

はじめに

2004年に結核予防法が改正され，これに伴い結核予防法施行令，結核予防法施行規則の一部も改正された[1)2)]．これにより，2005年度よりBCGワクチンの接種対象年齢が，改正前の「生後4歳に達するまで」（標準的接種期間は生後3カ月から1歳まで）から「出生時から生後6カ月に達するまで」に引き下げられた[2)]．改正法の施行直前に実施された全国BCGワクチン累積接種率調査によると，生後5カ月での全国BCGワクチン累積接種率は52.2%にすぎなかった[3)]．また，栃木県下の9市町における調査では，生後5カ月での累積接種率は0.0%から83.6%まで大きなばらつきがみられた[4)]．さらに，この改正には移行期間が設けられていなかったことなどから，急激な接種期間の短縮に各自治体が対応しきれずに，BCGワクチンの接種率が低下し，接種もれ者が増加することが懸念された[2)4)5)]．また，生下時からのBCGワクチン接種が容認されたため，未発見の先天性免疫不全症児にBCGワクチンを接種して全身性BCG感染症が発生することも危惧された[2)6)]．このため，法改正に伴う接種率及び接種月齢の変化を知る目的で，法改正後にBCGワクチン接種年齢に達した小児におけるBCGワクチンの全国累積接種率を調査した．

対象および方法

全国のBCGワクチン累積接種率調査は，すでに発表した方法により[3)]，2006年4月までに満1歳に達した小児5,000人を全国から無作為に抽出し，抽出された1歳児が居住する市区町村1,257カ所に調査協力依頼書，調査票，調査手順書を郵送して実施した．2005年の調査[4)]と同様に，当該市区町村の予防接種担当者に，標本として選出された小児がBCGワクチン接種を受けた月齢の調査を依頼し，回収された調査票をもとに全国のBCGワクチン累積接種率[7)]及び接種完了率[7)]を算定した．

成　績

1. 回収率

2006年6月に全国から無作為抽出された1,601カ所の市区町村に調査依頼状を発送した．2006年7月27日現在で，1,076カ所の自治体から回答が寄せられたので，市区町村数から算出した回収率は85.6%となった．無作為抽出した1歳児の数（標本数）は5,000名であり，うち4,368名分の記録が返送されたので，標本数から算出した回収率は87.4%となった．回収された記録のうち，BCGワクチンに関する記載がないものが12名分，個人情報保護などの理由で調査に協力できないと書かれたものが91名分，接種済みだが接種日が不明と記されたものが71名分あったため，これら174名分を除外し，BCG接種済みとの回答の4,127名分，未接種との回答の67名分，合計4,194名分（全標本数の83.9%）の記録を集計した．

2. BCGワクチン累積接種率

2006年4月までに満1歳に達した小児におけるBCGワクチン累積接種率は，生後2カ月では4.5±0.6%であったが，生後3カ月では57.5±1.5%，生後4

（平成18年10月6日受付）（平成19年1月18日受理）
別刷請求先：（〒113-8677）文京区本駒込3—18—22
東京都立駒込病院小児科　髙山　直秀

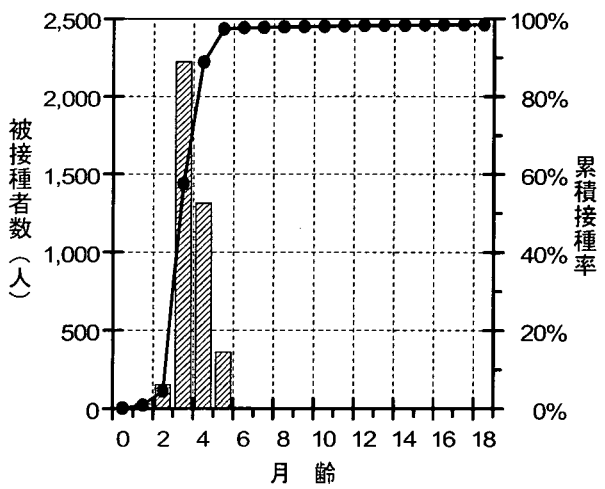

図1 BCGワクチンの月齢別接種者数及び累積接種率曲線

カ月では88.8±1.0%と急激に上昇し，生後5カ月では，すなわち生後6カ月に達するまでには97.4±0.5%に達した（図1）．2004年10月に3歳に達した小児を対象とした2005年の累積接種率調査では，生後3カ月で14.1±1.1%，生後5カ月では52.2±1.6%であった[4]ので，前回の調査結果と比較すると生後3〜5カ月での累積接種率が著しく向上していた．

3. BCGワクチン接種完了率

ある集団においてある年齢までにワクチン接種を完了した者の割合を示す完了率[6]をBCGワクチン接種対象月齢の0カ月から5カ月の乳児で算定すると，41.5%であり，この集団でBCGワクチン接種済み者は4割強にすぎなかった．しかし，この集団を月齢で2分すると，生後3カ月以前の乳児におけるBCGワクチン接種完了率は1.8%，生後3カ月から6カ月未満児（生後3〜5カ月児）における接種完了率は81.2%となり，生後3〜5カ月児では8割以上がBCGワクチン接種を済ませていることが判明した．さらに，生後6カ月から9カ月未満の乳児（生後6〜8カ月児）での完了率は97.8%，生後9カ月から12カ月未満の乳児（生後9〜11カ月児）では98.1%であった．

4. 月齢別BCGワクチン接種者の割合

BCGワクチンを生後0カ月で接種したとの回答は，集計母数4,194名中4名（0.1%），生後1カ月での接種が32名（0.8%），生後2カ月では153名（3.6%）と少数であった．一方，生後3カ月での接種は2,223名（53.0%）で最も多く，4カ月では1,313名（31.3%）と生後3カ月の約6割に減少し，5カ月ではさらに減少して359名（8.6%）であった．生後6カ月から9カ月で接種を受けた乳児がそれぞれ13，3，7，2名であり，生後10カ月，11カ月での接種者がそれぞれ3，5名であった．生後12カ月から17カ月までの接種者は合計10名であった（図1）．生後3，4，5カ月の3カ月間にBCG接種を受けた乳児は3,895名で集計母数の92.9%を占めた．

考 察

改正法実施後の全国BCGワクチン累積接種率は生後6カ月以前に約97%に達しており，「乳児期早期のBCG接種率を高め乳児の結核免疫力を強化する」という法改正の目的は達成されたと判断できる．接種時期に関しては，約93%の接種対象者が生後3カ月から6カ月に達する前までに接種を受けており，小児科学会などが全身性BCG感染症の発生を避けるために推奨している生後3〜6カ月という接種期間が受け入れられたものと思われる．制度の移行期間がなく，準備期間が短かったにもかかわらず，生後11カ月で98%という累積接種率が達成できたことは，各市区町村の予防接種関係者の多大な努力の賜と考えられる．一方，生後7カ月以降での接種者が少ないことは，接種漏れ者対策が軌道に乗っていないことを推測させる．生後1歳までの累積接種率が98%に達したとはいえ，BCGワクチンの対象者数を全国で120万人と仮定すると，未接種者の1歳児が全国で約2万4千人残されていることになるので，今後は接種漏れ者を早期に発見して接種を勧奨できる体制を整備する必要がある．また，すでに指摘されているように[4]，BCGワクチンの接種時期がジフテリア・百日咳・破傷風3種混合（DPT）ワクチン及びポリオ生ワクチンの1回目接種の時期と競合しているため，DPTワクチンやポリオ生ワクチンの接種率にどのような影響を与えているかを検証する必要がある．

文 献

1) 神谷 齊．結核予防法の改正の要点．小児科 2005；46：319—323.
2) 高松 勇．これからのBCG接種体制の注意点と今後の問題点．小児科 2005；46：324—332.
3) 高山直秀，崎山 弘，宮村達男，他．麻疹ワクチン及びポリオ生ワクチン累積接種率全国調査結果．感染症学誌 2005；79：7—12.
4) 高山直秀，崎山 弘，岡部信彦，他．結核予防法改正前の全国BCGワクチン累積接種率．日児誌 2006；110：14—16.
5) 高山直秀，手塚修一．結核予防法改正前の栃木県下市町におけるBCGワクチン累積接種率．小児科臨床 2006；59：1861—1865.
6) 高松 勇．BCG．小児科 2006；47：471—477.
7) 高山直秀，崎山 弘．各市区町村における麻疹ワクチン累積接種率調査結果．小児科臨床 2005；58：215—220.

Nationwide BCG Cumulative Vaccination Rates in Japan after the Revision of the Tuberculosis Prevention Law in 2005

Naohide Takayama[1], Hiroshi Sakiyama[2] and Nobuhiko Okabe[3]
[1]Tokyo Metropolitan Komagome Hospital
[2]Sakiyama Pediatric Clinic
[3]Infectious Disease Surveillance Center, National Institute of Infectious Diseases

The age for BCG vaccination in Japan was lowered to 0—5 months of age under the Tuberculosis Prevention Law at the beginning of the 2005 fiscal year. It was predicted that this revision would make the BCG-vaccination rate decrease, and there would be an increase in number of children who would miss the opportunity of vaccination. In order to know the changes in rates after the implementation, the nationwide cumulative vaccination rate (CVR) of the BCG vaccine among infants of the BCG-vaccination age was investigated. It revealed that the CVR after the revision law was executed had reached about 97% before 6 months of age. Furthermore, about 93% of infants received the BCG vaccine between 3—5 months of age. Remaining issues will be to prepare a system that could find infants missing the vaccination effectively and could encourage vaccination. It is also necessary to verify what influence the early BCG-vaccination might have to the vaccination rates of both diphtheria- pertussis- tetanus (DPT) vaccine and oral polio vaccine.

THE JOURNAL OF THE JAPANESE ASSOCIATION
FOR INFECTIOUS DISEASES
November, 2007, p675—680
0387—5911

妊婦における麻疹中和抗体価，HI抗体価，PA抗体価の相関と各測定法の発症予防レベル

[1] 東京都立駒込病院小児科，[2] 獨協医科大学産婦人科，[3] 千葉県衛生研究所感染疫学部

高山　直秀[1]　庄田亜紀子[2]　岡崎　隆行[2]
一戸　貞人[3]　斉加志津子[3]　稲葉　憲之[2]

感染症学雑誌　第81巻　第6号　別刷