Genotype analysis of varicella-zoster virus isolates from suburban Shanghai Municipal Province, China

Z. Sun,1 Y. Guo,1 M. Li2 and Z. Yao2

1Department of Dermatology, Fengxian Institute of Dermatosis Prevention, Shanghai, PR China
2Department of Dermatology, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, PR China

To determine the predominant genotype of the varicella-zoster virus (VZV) in suburban Shanghai Municipal Province, specimens were collected from the lesions of 95 outpatients clinically diagnosed with varicella or herpes zoster. Of these, 69 patients (72.6 %) were positive for VZV DNA. The 69 isolates were all genotyped as the genotype J1/clade 2. Based on sequencing of the 447 bp sequence in ORF22, 66 isolates were identified as genotype J/clade 2 strains and three were identified as type M2/clade 4 strains. To confirm the classification of these three strains, we determined the presence of 27 single-nucleotide polymorphisms (SNPs) and found that isolates 1270/1450 shared seven SNPs that differed from those of clade 2, in which three SNPs were unique to clade 3 and another three were unique to clade 4. Isolate 1456 had two markers of clade 4 that differed from clade 2. The phylogenetic tree showed that our isolates clustered primarily with clade 2 and that the three M2/J1 strains clustered between clades 2 and 4. It is likely that isolates 1270/1450/1456 may represent a new subclade of either clade 2 or 4, or some recombinant events. In addition, our isolates were WT strains. We also observed significant inter-strain variations.

INTRODUCTION

The genome of the varicella-zoster virus (VZV), a member of the family Herpesviridae and subfamily Alphaherpesvirinae, has been fully sequenced and found to consist of 125 kb of linear, dsDNA coding for 71 ORFs. VZV is the causative agent of varicella and herpes zoster (HZ) infections in humans. As a primary infection, varicella usually occurs in childhood, although it can be reactivated and manifest as HZ infection, or ‘shingles’, which typically occurs after the age of 50 years.

Various types of single-nucleotide polymorphism (SNP)-based analyses of the VZV genome have been performed using isolates collected throughout the world. By SNP analysis of ORFs 1, 21, 50 and 54, Barrett-Muir et al. (2003) identified four VZV genotypes based on the geographical locations of the isolates: A (Africa and Asia), J (Japan), B (Europe and North America) and C (Europe and North America), among which genotype B may represent a recombinant of genotypes C and A. By combined SNP analysis of a 447 bp region in ORFs 22 and 21, or 50, Loparev et al. (2004) identified five VZV genotypes (E1, E2, J, M1 and M2) and two provisional genotypes (M3 and M4). By SNP analysis of five glycoprotein genes and ORF 62, Faga et al. (2001) identified genotypes A, B, C and D, but their characteristics did not correspond to those reported by Barrett-Muir et al. (2003). Therefore, these three methods of VZV classification are not mutually compatible. With the development of whole-genome sequencing, a meeting was organized in London in July 2008, to create a common system for VZV classification and nomenclature. Based on the phylogeny of VZVs, five major clades (1–5) and two provisional clades (VI and VII) were proposed. For example, genotypes J and M2 in the Loparev system represent clades 2 and 4, respectively (Breuer et al., 2010). More recently, genotypes VIII and IX have been characterized with 454 technology as almost-complete genome sequences (Zell et al., 2012).

Previous studies to address the geographical variation in the epidemiology of VZV infection have reported that genotype J/clade 2 is the most prevalent VZV in temperate regions in China (Liu et al., 2009; Jiang et al., 2013). To investigate the distribution of VZV strains in suburban Shanghai Municipal Province, we performed genomic analyses of VZV isolates obtained from patients in suburban Shanghai from 2011 to 2013, using the methods of both Barrett-Muir et al. (2003) and Loparev et al. (2004).
To confirm the identity of putative clades, we analysed the isolates for the presence of the 27 SNPs proposed by Breuer et al. (2010). The findings of this study showed that genotype J/clade 2 was predominant in suburban Shanghai, while three isolates might represent a new subclade or recombination events. In addition, our isolates were WT strains. We also observed significant interstrain variations.

**METHODS**

**Patients and samples.** The study cohort included 95 outpatients clinically diagnosed with varicella or herpes zoster infection at the Department of Dermatology of the Foshan Institute of Dermatosis Prevention (Shanghai, China) between May 2011 and June 2013. Only 69 patient specimens were VZV-positive. The varicella cohort included 14 women and 10 men (mean age, 15 years; age range, 2–36 years) and the HZ cohort included 20 women and 25 men (mean age, 50 years; age range, 18–85 years). Of these patients, 59 resided in Shanghai Municipal Province (two in Shanghai City and 57 in suburban Shanghai) and 10 were from neighbouring provinces located in the middle-lower plains of the Yangtze River (five from Anhui Province, two from Jiangsu Province, two from Zhejiang Province and one from Jiangxi Province). Five varicella patients were vaccinated with Varilrix (GlaxoSmithKline Biologicals). Six varicella patients were adults. Two patients with disseminated herpes zoster were found to be immunocompetent. There was no unifying epidemiological factor among the outpatients and none were related or reported prior acquaintances with one another. After obtaining informed consent, specimens were collected from the herpes zoster/varicella lesions of patients for detection and analysis of VZV DNA.

**VZV genotype analysis.** DNA was extracted and purified from clinical samples using the EZNA Viral DNA kit (Omega Bio-Tek), according to the manufacturer’s instructions, and used as the templates for PCR amplification. The primers and protocol used for PCR amplification of VZV ORFs 1, 21, 22, 38, 54 and 62 were described elsewhere (Liu et al., 2009). After three isolates were identified as type M2 strains, we designed primers (Table S1, available in the online Supplementary Material) using Primer Premier software (www.premierbiosoft.com) to screen for 16 SNPs in ORFs 6, 12, 16, 17, 22, 35, 37, 55, 56, 60 and 66, to confirm whether these three isolates represented clade 4. The 27 SNPs examined, including 11 SNPs in ORFs 6, 12, 16, 17, 22, 35, 37, 55, 56, 60 and 66, were previously proposed by Breuer et al. (2010) for the classification of VZVs. PCR products were purified and sequenced using an ABI Prism 3730 automated sequencer (Applied Biosystems). Both DNA strands were sequenced whenever a novel or ambiguous nucleotide was found, as compared with the reference sequence. The nucleotide sequences were analysed and compared, using Geneious Pro version 4.8.4 software (www.geneious.com), with the lowest VZV-positive rate (data not shown). Sixty-nine patients (72.6%) tested positive for VZV DNA. We found a novel 9 bp in-frame (TACACCCAG) insertion between nucleotide positions 672 and 673 in ORF 1 of isolate 1262 (Fig. 1), relative to that of the Dumas reference strain. This insertion, which is transcribed from the negative strand, introduced three amino acid residues (L-G-V) into the putative transmembrane protein encoded by ORF 1. One T→A nucleotide substitution (SNP790) in ORF 1 was identified in six isolates (8.70%). Three consecutive T→C substitutions occurred at positions 789–791 in the majority of our isolates and the poKa strain, relative to the Dumas strain, which we used as a marker to differentiate between genotypes. A rare T→A transversion mutation was observed at position 790 in isolates 1273, 1427, 1428, 1429, 1453 and 1457. Because ORFs 1, 38 and 54 are transcribed from the negative strand, the codons complementary to TTG, CCG and ACG were transcribed/translated as CAA/Q, CGG/R and CGU/R in Dumas, poKa and the six isolates, respectively (Fig. 1). Other identified inter-strain variations included ORF21, which encodes the nucleocapsid protein, and ORFs 22 and 38, which encode tegument proteins. Non-synonymous nucleotide substitutions were identified at positions 33 722, 33 745, 37 990, 38 059 and 69 396, which resulted in an H→Q substitution in isolate 1269, a T→K substitution in isolates 1059 and 1435, a Q→R substitution in isolate 1264, a T→N substitution in isolates 1266 and 1445 and a P→A substitution in sample N18, respectively (Fig. 1). Synonymous nucleotide substitutions were identified at the following nucleotide positions in the isolates indicated: position 33 848 in isolate 1264; position 69 169 in isolates 1273 and 1453; position 69 304 in isolates 1266, 1437, 1444, 1445 and N1; position 69 322 in isolate 1448; position 69 424 in isolates 1270 and 1450; position 69 652 in isolates 1270 and 1450; position 95 094 in isolates 1274 and 1452 and position 95 436 in isolate N3 (Fig. 1). In addition, the PstI site in ORF 38 (A at position 69 349), the BglI site in ORF 54 (T at position 95 241) and the Smal and Nael sites (T at positions 106 262 and 107 252 in ORF 62) were usually used to distinguish between the vaccine and WT VZV strains. All of the isolates in the current study were PstI+ (A was positive) and BglI+ (T→C substitution was positive), as well as Smal− and Nael− (T→C substitution was negative).

**RESULTS**

To evaluate the VZV-positive rate of different types of lesions, we collected specimens from vesicles, maculopapules and scabs. Specimens collected from scabs had the lowest VZV-positive rate (data not shown). Sixty-nine patients (72.6%) tested positive for VZV DNA. We found a novel 9 bp in-frame (TACACCCAG) insertion between nucleotide positions 672 and 673 in ORF 1 of isolate 1262 (Fig. 1), relative to that of the Dumas reference strain. This insertion, which is transcribed from the negative strand, introduced three amino acid residues (L-G-V) into the putative transmembrane protein encoded by ORF 1. One T→A nucleotide substitution (SNP790) in ORF 1 was identified in six isolates (8.70%). Three consecutive T→C substitutions occurred at positions 789–791 in the majority of our isolates and the poKa strain, relative to the Dumas strain, which we used as a marker to differentiate between genotypes. A rare T→A transversion mutation was observed at position 790 in isolates 1273, 1427, 1428, 1429, 1453 and 1457. Because ORFs 1, 38 and 54 are transcribed from the negative strand, the codons complementary to TTG, CCG and ACG were transcribed/translated as CAA/Q, CGG/R and CGU/R in Dumas, poKa and the six isolates, respectively (Fig. 1). Other identified inter-strain variations included ORF21, which encodes the nucleocapsid protein, and ORFs 22 and 38, which encode tegument proteins. Non-synonymous nucleotide substitutions were identified at positions 33 722, 33 745, 37 990, 38 059 and 69 396, which resulted in an H→Q substitution in isolate 1269, a T→K substitution in isolates 1059 and 1435, a Q→R substitution in isolate 1264, a T→N substitution in isolates 1266 and 1445 and a P→A substitution in sample N18, respectively (Fig. 1). Synonymous nucleotide substitutions were identified at the following nucleotide positions in the isolates indicated: position 33 848 in isolate 1264; position 69 169 in isolates 1273 and 1453; position 69 304 in isolates 1266, 1437, 1444, 1445 and N1; position 69 322 in isolate 1448; position 69 424 in isolates 1270 and 1450; position 69 652 in isolates 1270 and 1450; position 95 094 in isolates 1274 and 1452 and position 95 436 in isolate N3 (Fig. 1). In addition, the PstI site in ORF 38 (A at position 69 349), the BglI site in ORF 54 (T at position 95 241) and the Smal and Nael sites (T at positions 106 262 and 107 252 in ORF 62) were usually used to distinguish between the vaccine and WT VZV strains. All of the isolates in the current study were PstI+ (A was positive) and BglI+ (T→C substitution was positive), as well as Smal− and Nael− (T→C substitution was negative).
Furthermore, the Varilrix SNPs (C) at positions 560, 703 and 603 were absent (Fig. 1). Therefore, all isolates in this study were determined to be WT strains (Fig. 1).

The sequences of the various VZV isolates were aligned with reference sequences and assigned to the corresponding genotypes (Fig. 1). The SNP profiles of ORF 22 indicated that 66 isolates were genotype J strains, whereas only isolates 1270, 1446 and 1450 were the M2 genotype. However, based on SNP analysis of ORFs 1, 21, 50 and 54, all 69 isolates were classified as genotype J1 according to the nomenclature scheme proposed by Barrett-Muir et al. (2003) (Fig. 1).

PCR amplification of the 27 SNPs proposed by Breuer et al. (2010) based on the common nomenclature revealed that isolates 1270 and 1450 had nucleotide A at position 5827 in ORF6, C at position 17 834 in ORF12, A at position 37 902 in ORF22, G at position 39 394 in ORF22 and T at position 68 254 in ORF37. Furthermore, each of these isolates was classified as genotype J1 according to the nomenclature scheme proposed by Barrett-Muir et al. (2003) (Fig. 1).
isolates had nucleotide A at position 69 424 and T at position 69 652 in ORF38, which were not observed in the 27-SNP profile. Each of these seven SNPs differed from clade 2. Three of these seven nucleotide changes (A at position 5827, T at position 68254 and A at position 69 424) were only present in clade 3. The T at position 69 652 was unique, which might represent a spontaneous mutation. An additional three are present in other clades, but none in clade 2. Meanwhile, the two isolates had eight SNPs that differed from clade 4 (A at position 5827, G at position 6850, G at position 23 294, G at position 24 578, G at position 38 019, T at position 68 254, T at position 98 825 and A at position 113 243) (Fig. 1). These isolates carried three markers of clade 4, in which A at positions 5827 and 3902 can also be found in clades 1, 3 and 5, and G at position 39 394, which was also found in clade 1.

Isolate 1446 differed at two positions, i.e. C at position 18 082 in ORF12 and A at position 39 394 in ORF22, which were markers of clade 4, but were also present in clades 1, 3 and 5. Nine of the 10 isolates had C at position 18 082, whereas isolate 1261 was more closely related to clade 2 (reference strain AB097933). The phylogenetic tree based on the 27 SNPs within the conserved sequences showed that isolates 1270 and 1450 clustered as an independent clade near clade 4, whereas isolate 1446 clustered as an independent clade between clades 2 and 4 (Fig. 2). Considering the above results, it is likely that isolates 1270/1450 might represent a branch of either clade 2 or 4, or some recombinant strains among clades 2, 3 and 4. Isolate 1446 might also be such a branch, but recombinant between clade 2 and 4.

**DISCUSSION**

The distribution of VZV strains is likely influenced by a variety of factors, including climate, virus–host interactions and immigration patterns. Despite the vast population of China and the significant potential for the negative impacts of VZV-related diseases, relatively few VZV isolates from China have been genotyped. A previous study found that three VZV isolates from southern China were type M2 strains (Loparev et al., 2004). The same study also examined three isolates from northern China, and found that they were all type E strains.

The samples in the present study were collected from suburban areas surrounding Shanghai City, one of the largest cities in the world, with a population of 30 million citizens and a floating population of more than 9 million migrants. Liu et al. (2009) identified 19 isolates from patients in Hefei, the capital city of Anhui Province in central-eastern China, as genotype J. Another study in Wuhan in central China showed that all of the 78 isolates examined were genotype J strains (Zhou et al., 2009). The cities of Hefei, Wuhan and Shanghai all lie on the flood plain of the Yangtze River at latitudes of approximately 31° N and each typically has a temperate climate.

![Fig. 2. Phylogenetic trees were reconstructed using the Geneious Tree Builder computational program included in the Geneious Pro software package, which employed the Tamura–Nei genetic distance model, a neighbour-joining tree-build method with 10 000 bootstrap replicates, and the CLUSTAL W plugin for alignment. The reference strains for each genotype were selected from reference (Breuer et al., 2010). Isolates 1270, 1446 and 1450 are shown in bold type. Twenty-seven SNPs within conserved sequences (4410 bp) in 10 isolates and reference strains were compared by phylogenetic analysis. Seven of the isolates (1260, 1261, 1265, 1431, 1436, 1437 and 1456) were randomly selected to represent genotype J/clade 2. The positions of the mutations in the 10 isolates are shown in Fig. 1. Bar, evolutionary distance.](image-url)
This study was the first to report VZV genotypes in suburban Shanghai, with Shanghai natives comprising the majority of our patients and approximately 15% of our patients consisting of migrants from neighbouring provinces. Our results showed that the VZV strains isolated from patients in suburban Shanghai were primarily genotype J/clade 2 strains, suggesting that the distribution of VZV in China has significant geographical features. The results of our study, combined with those from studies in Japan (Inoue et al., 2010), Korea (Kim et al., 2011) and Mongolia (Enkhsalkhan et al., 2010), indicate that the majority of WT VZV strains in north-east Asia are genotype J/clade 2 strains. However, a small fraction of type M2 isolates were identified according to the 447 bp method in our study. The reference strain for genotype M2 was the African strain DR (Loparev et al., 2007), which shared 99.9% nucleotide sequence identity with the Dumas strain. The DR strain and isolates 1270, 1446 and 1450 were assigned to genotype M2 based on the analysis of ORF 22 alone. However, each was classified as genotype J1 according to the criteria used by Barrett-Muir et al. (2003), the same genotype as the pOka and vOka strains. This method is incapable of differentiating between the M2 and J genotypes. In Australia and India, genotype M2 strains are predominant, and M2 strains have been identified to a much lesser extent in China (Loparev et al., 2004; Toi & Dwyer, 2010). To compare the three isolates in question with the five clades, we screened for the 27 SNPs proposed by Breuer et al. (2010) to confirm that they were M2/clade 4 strains. There were 10 SNPs across eight ORFs, which differentiated M2/clade 4 from J/clade 2, based on the minimum complement of SNPs required to identify a putative clade. We found that isolates 1270 and 1450 shared seven SNPs that differed from those of clade 2, whereas isolate 1456 had two SNPs that differed from clade 2. Three SNPs, including A at position 5827, T at position 68254 and A at position 69424, uniquely belonged to E2/clade 3. Three other changes were markers of M2/clade 4, but they were not unique, as they were also found in clades 1, 3 and 5. Finally, they had one spontaneous mutation (T at position 69652) that was not found in any clade. The phylogenetic tree analysis demonstrated that isolates 1270 and 1450 represented an independent clade near clade 4, and that isolate 1446 also represented an independent clade located between clade 2 and the novel clade represented by isolates 1270 and 1450. These findings suggested that isolates 1270/1450 might represent a branch of either clade 2 or clade 4, or some recombinant strains among clades 2, 3 and 4. Isolate 1446 might also be such branch, but recombinant between clades 2 and 4. Peters et al. (2006) suggested that clade 4 strains evolved through a recombination of the clade 1 and 2 strains. Currently five VZV clades are recognized. Clades 1 and 3 represent European strains, clade 2 represents Japanese strains and clade 5 Indian strains (Chow et al., 2013). Recently an analysis performed in the UK and Europe also showed that two recombinants might represent clade 5 variants due to immigration (Quinlivan et al., 2013). Therefore, it is possible that the influx of migrant workers, especially those from other countries, has resulted in the recombination of genotypes in the Shanghai area. Whole-genome sequencing should verify our findings.

The ORFs 38, 54 and 62 of the VZV isolates in our study were PstI+, BglII+, SmaI− and NaeI−, which is consistent with the findings of other studies of VZV strains in China. These four sites distinguished the WT and vaccine strains of VZV (Breuer, 2010; Schmid, 2010). However, PstI and BglII might not be reliable markers for this distinction. A previous study conducted in the US reported that the majority of circulating VZV strains lacked the BglII site in ORF 54, which is present in the vaccine strain. Virtually all of the circulating strains carried the PstI site in ORF 38, which is absent in the vOka strain (LRussa et al., 1998). Because the largest number of differences between the WT and vaccine strains of VZV occurred in ORF 62, the genome of strain vOka was mutated (T→C), creating SmaI, NaeI and BssHII sites (positions 106262, 107252 and 10811, respectively) in ORF 62. These introduced sites serve as vaccine-specific markers that allow the differentiation of various vaccine strains from the pOka parent strain and other WT VZV strains (Argaw et al., 2000).

Although the varicella vaccine has been distributed in China since 1998, vaccine-related chickenpox continues to be reported (Gan et al., 2011). The use of the three restriction sites to distinguish WT VZV infections from those of the vaccine strain is of epidemiological importance. In our current study, five isolates from vaccinated patients were genotyped as non-vaccine strains, suggesting that vaccination might not have been the cause of recent chickenpox or herpes zoster infections, which were previously attributed to vaccine infections. Hence, such VZV infections may have resulted from vaccine failure or depressed immunity (Schmidt-Chanasit & Sauerbrei, 2011).

Although most of the isolates in the current study were identified as genotype J strains, indicating a high degree of homology among VZV isolates in suburban Shanghai, significant variation was observed between the various isolates. A previous study of VZV genotypes in China reported a CGG insertion in ORF 1 (Jiang et al., 2013). We found a nine-base insertion in ORF 1 of isolate 1262. The SNPs in strain 37902 include five T→A single nucleotide substitutions, the combination of which is very rare because of the rarity of the pyrimidine to purine transversion. In ORFs 21, 22 and 38, which encode the nucleocapsid protein, we identified five missense mutations in seven isolates, only some of which have been reported elsewhere (Liu et al., 2009; Jiang et al., 2013). We also identified eight synonymous mutations in 16 isolates. Therefore, future studies are warranted to determine whether these mutations and SNPs affect protein function or alter immunogenicity.

In summary, the VZV genotype J1/J/clade 2 was predominant in suburban Shanghai. Although three isolates were
identified as J1/M2 genotype strains based on the methods of Barrett-Muir et al. (2003) and Loparev et al. (2004), they may represent a new subclade or recombinant events according to the 27-SNP profile. These recombinant strains were found only in Shanghai, reflecting an irregular distribution of VZV genotypes in China. China is a country of diverse geography and climate, with a variety of ethnicities. Future studies of the prevalence of VZV in additional cities and among different populations are warranted to clarify the distribution of VZV genotypes in China. In addition, although the sequencing of the 447 bp sequence in ORF22 can be used to rapidly and reliably discriminate mosaic types, use of the 27-SNP profile only confirmed five existing clades, which suggests that the VZV genotyping scheme should be based on the whole genome sequence. The addition of whole-genome sequencing for the characterization of novel isolates is also warranted to confirm our findings.

ACKNOWLEDGEMENTS

This study was financially supported by the Science and Technology Fund Bureau of Health of Fengxian, Shanghai, China (grant no. 2011-04) and the Shanghai Municipal Health and Family Planning Commission (grant no. 201440568).

REFERENCES


