Journal of Bacteriology and Virology 2014. Vol. 44, No. 4 p.326 – 335
http://dx.doi.org/10.4167/jbv.2014.44.4.326

Original Article

Characterization of the Repeat Sequences of Varicella-Zoster Virus

Youn Hee Won, Jong Ik Kim, Yu Young Kim and Chan Hee Lee*

Department of Microbiology, Chungbuk National University, Cheongju, Korea

Varicella-zoster virus (VZV) is a causative agent for shingles and herpes zoster. The genomes of VZV contain five reiteration (R) sequences and an origin of replication (ORI) sequences composed of tandem repeats whose numbers vary among different strains. Variation of the genome lengths among VZV strains could be attributed by the lengths of R sequences. There was a strong correlation between the lengths of VZV genome and R sequences, while variation of ORI did not contribute the variation of VZV genome length. The high G+C contents of The R sequences in ORF11, 14 and 22 influenced the codon usage of VZV in these ORFs. None of the most frequent 5 codons in R sequences was included in the top 5 most frequent codon in ORF11-14-22 or VZV genome, and vice versa.

Key Words: Varicella-zoster virus, Reiteration sequence, Codon usage

INTRODUCTION

Varicella-zoster virus (VZV) belongs to beta-herpesviridae and causes shingles during primary infection. Viruses establish latent infection in dorsal root ganglia and reactivation from latency often results in herpes zoster in older adults and in immunocompromised people. VZV was first isolated by T. Weller (1). VZV contains a linear double-stranded DNA genome with approximately 125 kbp. The VZV genome, like the genomes of herpes simplex virus (HSV) and cytomegalovirus (CMV) could be divided into four regions: terminal repeat long (TRL), unique long region (UL), internal repeat long (IRL), internal repeat short (IRS), unique short region (US), and terminal repeat short (TRS) (2). TRL and IRL, TRS and IRS are reverse complementary to each other (2, 3).

Coding region of the VZV genomes consists of 74 open reading frames (ORFs) and account for about 91% of the genome. VZV genome has been noted to be highly conserved (4~7). The herpesviruses in that the natural mutation rate was estimated to range between 10^{-6} and 10^{-7} substitution/site/year (8, 9), while that of herpes simplex virus type 1 ranged between 10^{-4} and 10^{-5} substitution/site/year (7, 9). Furthermore, the lengths of the VZV genomes are relatively constant compared to other herpesviruses. Many ORFs are invariant in their lengths among strains. Some ORFs with variable lengths contain reiteration (R) sequences.

VZV genome contains 5 R sequences, R1 to R5 (2, 10). R1, R2 and R3 are located in ORF 11, 14, and 22, respectively. R4 and R5 are located in noncoding region, R4 between ORF 62 and origin of replication (ORI) and R5 between ORF60 and ORF61. As noted for the Oka strains (10), the R sequences were found to vary not only

Received: October 2, 2014/ Revised: October 22, 2014/ Accepted: October 30, 2014
*Corresponding author: Chan Hee Lee, Ph.D. Department of Microbiology, College of Natural Sciences, Chungbuk National University, 52 Naesudong-Ro, Cheongju, Chungbuk 361-763, Korea.
Phone: +82-43-261-2304, Fax: +82-43-273-2451, e-mail: chlee@chungbuk.ac.kr
**This work was supported by the research grant of Chungbuk National University in 2012.

among the strains but within a strain (11, 12). Not much is known about the functions of R sequences.

Another variable region in the VZV genome is ORI. ORI is located between ORF62 and ORF63 in IRS, corresponding to 110,166~110,263 bp of the reference strain Dumas (2). A basic motif of tandem TA repeats followed by tandem GA repeats is flanked by 30 bp at 5' and 22 bp at 3'. The numbers of TA and GA repeats are varied among the VZV strains (11). Reverse complementary sequence of ORI is repeated in TRS, between ORF70 and ORF71. In this study, variable regions such as R and ORI sequences were characterized using the complete genomic DNA sequences of 50 VZV strains from GenBank database.

MATERIALS AND METHODS

VZV genome DNA sequences

Complete genomic DNA sequences from 50 VZV strains were retrieved from NCBI GenBank database. The 50 strains included 3 clinical strains isolated from Korean patients, YC01, YC02 and YC03. Five strains were vaccine strains: vOka, VarilRix, 1002/2008 and VariVax were derived from Japanese pOka strain and Suduvax was derived from Korean MAV/06 strain. The 50 VZV strains analyzed in this study and their GenBank accession numbers are listed in Table 1.

Identification of reiteration (R) and replication origin (ORI) region

Nucleotide sequences of the full genomes of 50 VZV strains were multiple aligned with ClustalW (ver.2.1). R and ORI regions were identified first from the reference strain Dumas (NC001348); R1 (13,937-14,242), R2 (20,692-21,017), R3 (41,454-41,519), R4a (109,762-109,907), R4b (119,990-120,135), R5 (102,020-102,219), ORI (110,166-110,263). R and ORI regions of the other strains were determined as the corresponding regions to Dumas strains and excised from the genome sequences.

Analysis of codon usage

GenBank files in .gb file format were downloaded and used as input files to SeqMan (ver. 1.0) program developed by Virology Lab, Chungbuk National University. Address files for the coding DNA sequence (CDS) in .txt file format and DNA sequence files in .fst file format were extracted with SeqMan program. Fst files containing DNA sequences of each CDS were obtained and used as input files to CodonW program (ver.1.4.2, http://codonw.sourceforge.net/) in order to get the frequencies of the codons. Universal genetic code was used as the default genetic code. Resulting .blk file contained the number of 64 codons for each CDS. Frequency of the each codon was calculated by dividing the number of each codon by the total number of 64 codons. Codon frequencies of the R region were determined from the excised R1, R2 and R3 as described above. Codon frequencies of genome were obtained from the concatenated CDS.

RESULTS AND DISCUSSION

Reiteration (R) sequences and genome length polymorphism

R1 was composed of a combination of multiple sequences of 18 bp elements (consensus: GGACGCGATCGACGACGA) and 15 bp elements (consensus: GGGAGAGGCGGAGGA). SuduVax differed from Oka vaccine strains in the length of ORF11 due to an additional copy of 15 bp element in SuduVax. R1s of the 3 Korean strains YC01, YC02 and YC03 strains were same as R1 of Japanese pOka.

R2 consisted of multiple 42 bp elements (ACCTCGGCCGCTT/aCCCGAAAG/taCCCGATCCCGCCGTCGCGCCC: lower case for minor deviation of the consensus) followed by one copy of its partial 32 bp element. Korean strains including SuduVax, YC01, YC02 and YC03 contained 7 copies of 42 bp element as Japanese strains did. Although LAX1 was considered to be clade 2 strain (13, 14) which includes Korean and Japanese strains, it contained 4 copies of 42 bp element. Thus, among the 11 clade 2 strains, LAX1 strain was the only one with different R2 pattern from the others.

R3 sequences consisted of repeating copies of 9 bp elements and are located at the 3' half in ORF22. The

Table 1. Reiteration (R) Sequences of VZV strains

Strain	GenBank accession number	Length (bp)								
		R1	R2	R3	R4a	R4b	R5	R^1	G^2	G-R^3
Dumas	NC001348	303	326	72	146	146	200	1,193	124,884	123,691
M2DR	DQ452050	303	284	27	146	146	200	1,106	124,770	123,664
CA123	DQ457052	258	284	36	146	146	200	1,070	124,771	123,701
SD	DQ479953	258	326	72	281	281	200	1,418	125,087	123,669
Kel	DQ479954	321	326	72	389	389	200	1,697	125,374	123,677
11	DQ479955	225	326	225	362	362	200	1,700	125,370	123,670
22	DQ479956	195	284	72	200	200	200	1,151	124,868	123,717
03-500	DQ479957	225	284	657	92	92	200	1,550	125,239	123,689
36	DQ479958	321	326	72	227	227	200	1,373	125,030	123,657
49	DQ479959	321	410	54	200	200	200	1,385	125,041	123,656
8	DQ479960	258	620	36	335	335	200	1,784	125,451	123,667
32p5	DQ479961	243	326	54	227	227	200	1,277	124,945	123,668
32p22	DQ479962	291	326	36	281	281	200	1,415	125,084	123,669
32p72	DQ479963	291	326	117	281	281	200	1,496	125,169	123,673
NH29_3	DQ674250	258	326	54	146	146	200	1,130	124,811	123,681
SVETA	EU154348	243	326	72	146	146	200	1,133	124,813	123,680
MSP	AY548170	258	368	90	146	146	200	1,208	124,883	123,675
BC	AY548171	258	326	612	200	200	200	1,796	125,459	123,663
HJ0	AJ871403	225	326	72	146	254	200	1,223	124,928	123,705
3/2005	JN704700	294	206	72	146	146	200	1,064	124,756	123,692
52/2007	JN704701	258	326	54	173	146	200	1,157	124,816	123,659
134/2005	JN704706	303	326	72	146	146	200	1,193	124,846	123,653
243/2000	JN704690	300	290	72	146	146	200	1,154	124,845	123,691
405/2007	JN704702	207	227	72	146	146	200	998	124,697	123,699
413/2000	JN704704	291	305	72	146	146	200	1,160	124,838	123,678
432/2008	JN704695	285	326	72	146	146	200	1,175	124,867	123,692
446/2007	JN704707	291	326	72	146	146	200	1,181	124,861	123,680
457/2008	JN704710	258	200	27	146	146	200	977	124,472	123,495
551/2005	JN704692	273	176	72	146	146	200	1,013	124,688	123,675
667/2005	JN704693	303	326	72	146	146	200	1,193	124,884	123,691
875/2004	JN704705	297	299	72	146	146	200	1,160	124,844	123,684
925/2008	JN704696	267	326	72	146	146	200	1,157	124,848	123,691
1219/2007	JN704703	201	152	72	146	146	200	917	124,617	123,700
1256/2004	JN704691	258	299	18	146	146	200	1,067	124,757	123,690
1483/2005	JN704709	198	263	72	146	146	200	1,025	124,718	123,693
1805/2007	JN704708	297	326	45	146	146	200	1,160	124,848	123,688

Table 1. Continued

Strain	GenBank accession number	Length (bp)								
		R1	R2	R3	R4a	R4b	R5	R[1]	G[2]	G-R[3]
1883/2007	JN704694	270	326	72	146	146	200	1,160	124,851	123,691
2308/2003	JN704699	300	296	72	146	146	200	1,160	124,847	123,687
pOka	AB097933	225	326	36	281	281	312	1,461	125,125	123,664
1003/2008	JN704698	210	326	72	146	146	200	1,100	124,772	123,672
LAX1	JQ972914	273	200	45	200	200	200	1,118	124,763	123,645
Ellen	JQ972913	225	284	54	173	173	200	1,109	124,783	123,674
YC01	KJ808816	225	326	36	281	281	312	1,461	125,144	123,683
YC02	KJ767491	225	326	36	281	281	312	1,461	125,150	123,689
YC03	KJ767492	225	326	36	281	281	312	1,461	125,162	123,701
vOka	AB097932	258	326	27	254	254	312	1,431	125,078	123,647
VarilRix	DQ008354	258	326	72	146	146	200	1,148	124,821	123,673
1002/2008	JN704697	237	323	72	146	146	200	1,124	124,814	123,690
VariVax	DQ008355	258	326	72	146	146	200	1,148	124,815	123,667
Suduvax	JF306641	273	326	99	92	92	200	1,082	124,758	123,676
Average		261.9	310.2	87.7	186.5	188.1	211.2	1,245.60	124,921	123,676
StDev[4]		34.5	65.2	116.9	68.6	69.2	33.9	210.2	212.2	30.3
CV[5]		0.132	0.21	1.333	0.368	0.368	0.161	0.169	0.0017	0.00024

[1] Sum of R sequences
[2] Length of the genome
[3] Length of the genome after deletion of R sequences
[4] Standard deviation
[5] Coefficient of variation, Standard deviation/Average

consensus sequence of the 9 bp element was GC/tCCGC/tG/cCA/g (lower case stands for minor deviation of the consensus). The number of repetition (n) varied tremendously among the strains as noted by Tyler *et al*. (11). In fact, the CV (coefficient of variation) value of R3 was calculated to 1.33, largest among the CVs of 5 R sequences (Table 1). The strain 03~500 contained the most repetition with n = 73, while vOka contained the least repetition with n = 3. SuduVax contains 11 copies of the 9 bp elements while both VarilRix and Varivax contain 8 copies.

R4 and R5 were located in non-coding region. R4 was repeating 27 bp elements (consensus: CCCCGCCGATG-GGGAGGGGGCGCGGTA) followed by its partial 11 bp element. One copy of R4 (R4a) was located in IRS between ORFs 62 and 63 and its reverse complement copy (R4b) was located in TRS between ORFs 70 and 71. Although the length of R4a was supposed to be same as that of R4b, the lengths of R4a and R4b were different in the strain HJ0 and 52/2007 strains. R4b of the strain HJ0 was 108 bp longer than R4a due to the additional 4 copies of the 27 bp elements in R4b. R4b of the strain 52/2007 was shorter than R4a by 1 copy of the 27 bp element. Clinical strains of the clade 2 contained 10 copies of the 27 bp elements, while vaccine strains in the clade 2 varied in the number of the 27 bp elements. VarilRix and VariVax contained 5 repeating units of the 27 bp element, although their parent vaccine strain vOka contained 9 copies of the 27 bp elements. SuduVax contained 3 copies of the 27 bp element.

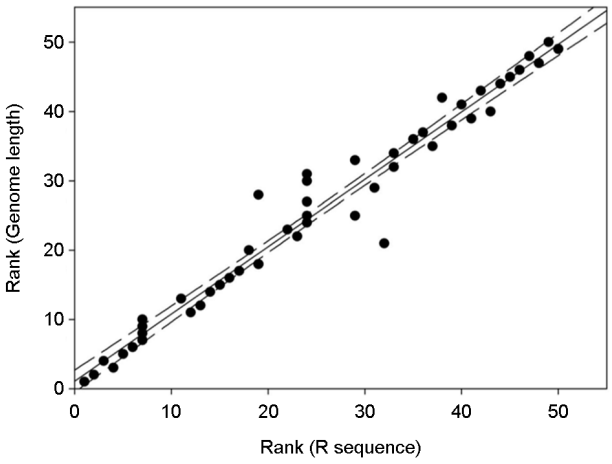

Figure 1. Correlation between the lengths of VZV genome and R sequences. The ranks of genome length among the VZV strains were plotted against the ranks of R sequences. Coefficient of correlation (r^2) of the linear regression line was calculated to be 0.9625. Solid line, simple linear regression line; dotted line, 95% confidence interval. y = 0.972x + 1.082

R5 was located between ORFs 60 and 61 and composed of 88 bp elements with 24 bp elements in between. The two elements were not related as observed in other R sequences. There were only two types of R sequences. Three Korean clinical strains YC01, YC02 and YC03 as well as Japanese pOka and vOka contained three 88 bp elements and two 24 bp elements. SuduVax and other vaccine strains contained two 88 bp elements and one 24 bp element as in most other strains.

Examination and comparison of the lengths of R and genome sequences among the VZV strains revealed that the variability of the VZV genome length was mostly due to the variable R lengths. The strain BC contains the longest genome and also the longest R sequence, while the strain 457/2008 contains the shortest genome and the second shorted R sequence (Table 1). The difference between the longest genome and shortest genome (987 bp) is similar to the difference between the longest and shortest R sequences (879 bp). If the R sequences were removed from the genome, the variability of the genome length was reduced as shown by the change in CV (coefficient of variation). CV was calculated to be 0.0017 at whole genome level and dropped to 0.00024 if the R sequences were removed from the genome (Table 1). In fact there was a linear relationship between the lengths of VZV genome and R sequences with coefficient of correlation (r^2) of 0.9625 when the order of the genome lengths were plotted against the order of the R lengths (Fig. 1).

Presence of tandem repeats in several parts of the VZV genome was proposed by restriction fragment length polymorphism (15) and later precisely mapped to VZV genome sequence (2). To date, 5 reiteration (R) sequences have been identified and there exist a certain degree of strain-to-strain variability of the lengths of the R sequences due to the variable number of the repeating elements (11). The lengths of the R sequences were not conserved even among the strains with same lineage, such as 4 Oka-derived strains (pOka, vOka, VarilRix, and VariVax) and 3 strain 32's with different passages (32p5, 32p22, and 32p72). Although highly conserved compared to the genomes of other herpes viruses, the lengths of the VZV genomes exhibit a small degree of variability. This genome-length polymorphism could be accounted for the variable lengths of R sequences among the VZV strains.

Origin of replication

The origin of replication (ORI) sequence of VZV was located between ORF62 and ORF63. The ORI sequences are duplicated between ORF70 and ORF71 in exactly reverse complementary form, except for the strains HJ0, 52/2007, YC03 and VarilRix. Lengths of ORI sequences varied from 80 bp in the strain 36 and 49 to 108 bp in HJ0 due to different number of tandem repeats of dinucleotide TA and GA at the centre of the ORI sequences (Table 2). The lengths of the ORI sequences of the Korean strains were found to be 104 bp unanimously, although the numbers of tandem TA repeats and tandem GA repeats were different. Among the other strains only pOka contained 104 bp ORI sequence. Although the lengths of ORI varied among the VZV strains, many strains shared same lengths of ORI. The lengths of the ORI sequences did not have correlation with the lengths of the genomes ($r^2 = 0.082$). Thus, variation of ORI did not contribute the variation of VZV genome length.

Table 2. Origin of replication (ORI) sequences of VZV strains

Strain	Start	End	Length	(TA)n[1]	(GA)n[2]
Dumas	110,166	110,263	98	16	7
M2DR	110,052	110,149	98	16	7
CA123	110,050	110,147	98	16	7
SD	110,250	110,337	88	12	6
Kel	110,431	110,514	84	10	6
11	110,455	110,538	84	9	7
22	110,075	110,180	106	14	13
03-500	110,580	110,667	88	12	6
36	110,262	110,341	80	8	6
49	110,300	110,379	80	8	6
8	110,541	110,632	92	14	6
32p5	110,171	110,254	84	10	6
32p22	110,256	110,339	84	10	6
32p72	110,335	110,420	86	11	6
NH29_3	110,094	110,191	98	16	7
SVETA	110,101	110,194	94	14	7
MSP	110,176	110,265	90	13	6
BC	110,708	110,795	88	12	6
HJ0	110,085	110,192	108	12	16
3/2005	110,035	110,132	98	16	7
52/2007	110,118	110,201	84	11	5
134/2005	110,151	110,234	84	9	7
243/2000	110,127	110,224	98	16	7
405/2007	109,975	110,072	98	13	10
413/2000	110,119	110,214	96	16	6
432/2008	110,149	110,246	98	16	7
446/2007	110,142	110,237	96	16	6
457/2008	109,848	109,945	98	16	7
551/2005	109,986	110,075	90	13	6
667/2005	110,166	110,263	98	16	7
875/2004	110,121	110,220	100	10	14
925/2008	110,130	110,227	98	16	7
1219/2007	109,894	109,991	98	16	7
1256/2004	110,039	110,136	98	16	7
1483/2005	109,993	110,090	98	16	7
1805/2007	110,123	110,222	100	16	8
1883/2007	110,133	110,230	98	16	7
2308/2003	110,128	110,223	96	16	6
pOka	110,268	110,371	104	12	14
1003/2008	110,049	110,148	100	13	11
LAX1	110,015	110,098	84	8	8
Ellen	110,060	110,145	86	9	8
YC01	110,277	110,380	104	12	14
YC02	110,279	110,382	104	12	14
YC03	110,282	110,385	104	10	16
vOka	110,264	110,359	96	9	13
VarilRix	110,096	110,191	96	9	13
1002/2008	110,091	110,190	100	12	12
VariVax	110,093	110,190	98	9	14
Suduvax	110,079	110,182	104	12	14

[1] The number of tandem TA repeats
[2] The number of tandem CA repeats

Characteristic codon usage in VZV genome and R sequences

Examination of the R sequences revealed several characteristic features. At nucleotide level, R sequences were highly GC-rich. While the average G+C content of VZV genomes was 46.05%, the average G+C content was calculated to be 70.41 ± 0.56% for R1, 75.24 ± 0.64% for R2, 84.46 ± 2.59% for R3, 81.5 ± 0% for R4 and 45.3 ± 0.41% for R5 (Table 3). Three of the R sequences, R1, R2 and R3, are located in ORF. Thus it was expected that the high G+C content might affect biased codon usage in R sequences. In order to test this possibility, the frequencies of the 64 codons were determined from R1, R2, R3, and concatenated R1-R2-R3. For comparison, codon frequencies were also obtained from concatenated ORF11-14-22 where R sequences are located. The codon frequencies of the entire genome were also obtained from the concatenated all ORFs.

The 3 most prevalent codons in R sequences were found to be GAG (for Glu, 17.54%), followed by CCC (for Pro, 13.97%), and GCG (for Ala, 12.84%) (Table 4, bolded and

Table 3. G+C % in reiteration sequences

R1	R2	R3	R4	R5	Genome
70.41±0.56	75.24±0.64	84.46±2.59	81.5±0	45.3±0.41	46.05±0.068

Table 4. Codon usage in VZV

Amino acid	Codon	Frequency (%)					
		R1	R2	R3	R1-R2-R3	ORF 11-14-22	Genome
Average±SD		82±12	103±22	30±39	216±46	4162±46	37884±86
Ala	GCA	0	0.98	0	0.47	2.31	2.04
Ala	GCC	0	**13.76**	0	6.66	1.98	2.13
Ala	GCG	**18.88**	6.32	**19.08**	**12.84**	2.00	1.85
Ala	GCU	0	7.37	0	3.56	1.84	1.54
Arg	AGA	0.03	0	0	0.01	0.83	1
Arg	AGG	0.18	0	0	0.09	0.45	0.59
Arg	CGA	0.03	7.35	0	3.57	1.08	1.18
Arg	CGC	0	0	0	0.00	0.87	1.19
Arg	CGG	0.05	0	2.27	0.28	0.99	1.11
Arg	CGU	0	0	0	0.00	1.38	1.48
Asn	AAC	0	0	0	0.00	1.51	1.89
Asn	AAU	0	0.77	0	0.38	2.31	2.17
Asp	GAC	**17.67**	0	0	6.88	2.23	2.16
Asp	GAU	0.35	7.36	0	3.69	**3.58**	**3.16**
Cys	UGC	0	0	0	0.00	0.30	0.64
Cys	UGU	0	0	0	0.00	1.09	1.49
Gln	CAA	0	0	0	0.00	2.43	2.14
Gln	CAG	0	0	**29.58**	3.78	1.26	1.37
Glu	GAA	0	0	0	0.00	**3.39**	**3.17**
Glu	GAG	**44.89**	0.02	0	**17.54**	2.43	1.95
Gly	GGA	11.98	0	0	4.67	1.74	2.27
Gly	GGC	0	0	0	0.00	0.63	0.91
Gly	GGG	0	0	0	0.00	1.14	1.56
Gly	GGU	0	0	0	0.00	1.28	1.43
His	CAC	0	0	0	0.00	0.83	1.08
His	CAU	0	0	0	0.00	1.41	1.54
Ile	AUA	0	0	0	0.00	2.21	1.92
Ile	AUC	4.69	0.03	0	1.81	0.87	1.03
Ile	AUU	0.79	0	0	0.34	2.91	2.58

Table 4. Continued

Amino acid	Codon	Frequency (%)					
		R1	R2	R3	R1-R2-R3	ORF 11-14-22	Genome
Average±SD		82±12	103±22	30±39	216±46	4162±46	37884±86
Leu	CUA	0	0	0	0.00	1.01	1.12
Leu	CUC	0	0.02	0	0.01	0.66	0.71
Leu	CUG	0	0	0	0.00	0.85	1.23
Leu	CUU	0	0	0	0.00	1.86	1.73
Leu	UUA	0	0	0	0.00	**3.39**	3.08
Leu	UUG	0.03	0	0	0.01	1.74	1.72
Lys	AAA	0	0.09	0	0.05	2.48	2.44
Lys	AAG	0	6.37	0	3.07	1.00	1.06
Met	AUG	0.02	0	4.89	0.48	1.93	2.09
Phe	UUC	0	0	0	0.00	0.47	0.68
Phe	UUU	0	0	0	0.00	2.88	**3.38**
Pro	CCA	0	0	0	0.00	1.64	1.75
Pro	CCC	0	**20.99**	**29.87**	**13.97**	2.30	1.76
Pro	CCG	0	0	0	0.00	0.87	1.43
Pro	CCU	0	0	0	0.00	1.17	1.15
Ser	AGC	0	0	0	0.00	0.54	0.9
Ser	AGU	0	0	0	0.00	1.10	0.99
Ser	UCA	0	0	0	0.00	1.11	1.32
Ser	UCC	0	3.5	0	1.70	1.49	1.46
Ser	UCG	0	7.44	0	3.60	1.17	1.23
Ser	UCU	0	0	0	0.00	1.79	1.49
TER	UAA	0	0	0	0.00	0.07	0.13
TER	UAG	0	0	0	0.00	0.00	0.03
TER	UGA	0	0	0	0.00	0.00	0.03
Thr	ACA	0	0	0	0.00	2.85	2.37
Thr	ACC	0	**11.26**	1.64	5.66	2.20	1.99
Thr	ACG	0.42	0.02	0	0.18	1.72	1.71
Thr	ACU	0	0	0	0.00	1.29	1.2
Trp	UGG	0	0	0	0.00	1.04	1.07
Tyr	UAC	0	0	0	0.00	0.99	1.29
Tyr	UAU	0	0	0	0.00	2.13	2.08
Val	GUA	0	0	0	0.00	1.95	1.89
Val	GUC	0	6.36	12.39	4.59	1.06	0.91
Val	GUG	0	0	0.29	0.03	1.36	1.65
Val	GUU	0	0	0	0.00	2.64	2.37

underlined). GAG was the most frequent codon in R1 and CCC was the most frequent codon in R2 and R3. The 3 most frequent codons accounted for 44.35% of all possible codons in R sequences, but only 6.73% in ORFs 11-14-22 and 5.56% in VZV genome (Table 4). On the other hands, 33 codons plus 3 termination codons were not present at all in the R sequences. None of the most frequent 3 codons in R sequences was included in the top 3 most frequent codon in ORF11-14-22 or VZV genome, and vice versa. Furthermore, 2 (GAA, UUA) of the top 3 most prevalent codons in ORF11-14-22 or 2 (GAA, UUU) of the top 3 most prevalent codons in genome were not present in R sequences. Thus, the high G+C content of the R sequences could account for the biased codon usage in the R sequences.

R1 is located in ORF11 which is a structural component of the tegument and known to be a determinant of VZV virulence in differential epidemical cells in their tissue microenvironment *in vivo* (16). However, the variable sign of the R1 in ORF11 did not affect VZV replication *in vitro* and in human skin *in vivo* (17).

R2 is located in ORF14 which coded for glycoprotein (gp) C. Yoshida *et al*. (18) observed that pOka and Oka-derived vaccine strains all contained 7 42-bp repeat unit and the number of 7 42-bp repeat unit is suggested to be relatively stable during viral replication and latency in human (19). All Korean strains including clinical strains and vaccine strain contained 7 42-bp repeat unit.

R3 is located in ORF22 coding for the largest VZV protein, the inner tegument protein (20, 21). ORF 22 is conserved throughout the herpesviridae. In other herpesviridae, pUL36 which is equivalent of VZV ORF22, contain two large stretcher of proline- and alanine- rich sequence. In VZV, 9-bp repeat unit G/CCC/GCG/CA contained proline (CCC) and alanine (GCG) residues which provided proline- and alanine- rich sequence in ORF22.

REFERENCES

1) Weller TH, Witton HM, Bell EJ. The etiologic agents of varicella and herpes zoster; isolation, propagation, and cultural characteristics *in vitro*. J Exp Med 1958; 108:843-68.
2) Davison AJ, Scott JE. The complete DNA sequence of varicella-zoster virus. J Gen Virol 1986;67:1759-816.
3) Arvin AM. Varicella-zoster virus. Clin Microbiol Rev 1996;9:361-81.
4) Norberg P, Liljeqvist JA, Bergström T, Sammons S, Schmid DS, Loparev VN. Complete-genome phylogenetic approach to varicella-zoster virus evolution: genetic divergence and evidence for recombination. J Virol 2006;80:9569-76.
5) Peters GA, Tyler SD, Grose C, Severini A, Gray MJ, Upton C, *et al*. A full-genome phylogenetic analysis of varicella-zoster virus reveals a novel origin of replication-based genotyping scheme and evidence of recombination between major circulating clades. J Virol 2006;80:9850-60.
6) Loparev VN, Rubtcova EN, Bostik V, Tzaneva V, Sauerbrei A, Robo A, *et al*. Distribution of varicella-zoster virus (VZV) wild-type genotypes in northern and southern Europe: evidence for high conservation of circulating genotypes. Virology 2009;383:216-25.
7) Tang JW. Phylogenetic studies of frequently diagnostically sampled herpesviruses--possibilities for clinical applications? Infect Genet Evol 2013;18:379-86.
8) Barrett-Muir W, Scott FT, Aaby P, John J, Matondo P, Chaudhry QL, *et al*. Genetic variation of varicella-zoster virus: evidence for geographical separation of strains. J Med Virol 2003;70:S42-7.
9) Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol 2010;27:2038-51.
10) Gomi Y, Sunamachi H, Mori Y, Nagaike K, Takahashi M, Yamanishi K. Comparison of the complete DNA sequences of the Oka varicella vaccine and its parental virus. J Virol 2002;76:11447-59.
11) Tyler SD, Peters GA, Grose C, Severini A, Gray MJ, Upton C, *et al*. Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. Virology 2007;359:447-58.
12) Cohen JI. The varicella-zoster virus genome. Curr Top Microbiol Immunol 2010;342:1-14.

13) Peters GA, Tyler SD, Carpenter JE, Jackson W, Mori Y, Arvin AM, *et al*. The attenuated genotype of varicella-zoster virus includes an ORF0 transitional stop codon mutation. J Virol 2012;86:10695-703.
14) Chow VT, Tipples GA, Grose C. Bioinformatics of varicella-zoster virus: single nucleotide polymorphisms define clades and attenuated vaccine genotypes. Infect Genet Evol 2013;18:351-6.
15) Straus SE, Hay J, Smith H, Owens J. Genome differences among varicella-zoster virus isolates. J Gen Virol 1983; 64:1031-41.
16) Che X, Oliver SL, Reichelt M, Sommer MH, Haas J, Roviš TL, *et al*. ORF11 protein interacts with the ORF9 essential tegument protein in varicella-zoster virus infection. J Virol 2013;87:5106-17.
17) Che X, Oliver SL, Sommer MH, Rajamani J, Reichelt M, Arvin AM. Identification and functional characterization of the Varicella zoster virus ORF11 gene product. Virology 2011;412:156-66.
18) Yoshida M, Tamura T, Miyasaka K, Shimizu A, Ohashi N, Itoh M. Analysis of numbers of repeated units in R2 region among varicella-zoster virus strains. J Dermatol Sci 2003;31:129-33.
19) Sauerbrei A, Zell R, Wutzler P. Analysis of repeat units in the R2 region among different Oka varicella-zoster virus vaccine strains and wild-type strains in Germany. Intervirology 2007;50:40-4.
20) Sandbaumhüter M, Döhner K, Schipke J, Binz A, Pohlmann A, Sodeik B, *et al*. Cytosolic herpes simplex virus capsids not only require binding inner tegument protein pUL36 but also pUL37 for active transport prior to secondary envelopment. Cell Microbiol 2013; 15:248-69.
21) Böttcher S, Klupp BG, Granzow H, Fuchs W, Michael K, Mettenleiter TC. Identification of a 709-amino-acid internal nonessential region within the essential conserved tegument protein (p)UL36 of pseudorabies virus. J Virol 2006;80:9910-5.