Comparison of the whole genome sequence of an Oka varicella vaccine from China with other Oka vaccine strains reveals sites putatively critical for vaccine efficacy.

Running title (51 characters):

WGS analysis of an Oka varicella vaccine from China

Qiuhua Wu, Pierre Rivailler, Songtao Xu and Wenbo Xu

"NHC Key Laboratory of Medical Virology and Viral Diseases (National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention), WHO WPRO Regional Reference Measles/Rubella Laboratory, Beijing 102206, People's Republic of China.

#Corresponding authors: Songtao Xu, xsttz886@sina.com; Wenbo Xu, wenbo_xu1@aliyun.com
ABSTRACT

Varicella-Zoster Virus (VZV) infection results in varicella mostly in children. Reactivation of the virus causes Herpes Zoster (HZ) mostly in adults. A live attenuated vaccine (vOka-Biken hereafter) was originally derived from the parental strain pOka. Several live attenuated vaccines based on the Oka strain are currently available worldwide. In China, varicella vaccines have been licensed by 4 manufacturers. In this study, we analyze the Whole Genome Sequence (WGS) of vOka-BK produced by Changchun BCHT Biotechnology also known as Baike. vOka-BK WGS was compared against the genomic sequence of 4 other Oka strains: pOka, vOka-Biken, vOka-Varilrix from GlaxoSmithKline (GSK) and vOka-Varivax from Merck & Co.. A previous study identified 137 Single Nucleotide Polymorphisms (SNPs) shared by all vOkas. The current analysis used this data as a reference to compare with vOka-BK WGS and focused on 54 SNPs located in the unique regions of the genome. Twenty-eight non-synonymous substitutions were identified, ORF62 and ORF55 featuring the most amino acid changes with 9 and 3 respectively. Among the 54 SNPs, 10 had a different mutation profile in vOka-BK compared to the other 3 vaccines. A comparison with the clade 3 strain Ellen, known to be attenuated, identified 3 shared amino acid changes, *130R in ORF0, R958G and S628G in ORF62. This analysis provides the first comparison of a Chinese varicella vaccine to the other vaccines available worldwide and identifies sites potentially critical for VZV vaccine efficacy.

IMPORTANCE
Varicella, also known as chickenpox, is a highly contagious disease, caused by Varicella Zoster Virus (VZV). Varicella is a common childhood disease that can be prevented by a live attenuated vaccine. The first available vaccine was derived from the parental Oka strain in Japan in 1974. Several live attenuated vaccines based on the Oka strain are currently available worldwide. Among the 4 vaccines produced in China, the vaccine manufactured by Changchun BCHT Biotechnology also known as Baike, has been reported to be very efficacious. Comparative genomic analysis of the Baike vaccine with other Oka vaccine strains identified sites that might be involved in vaccine efficacy as well as important for the biology of the virus.

INTRODUCTION

Varicella-zoster virus (VZV, Human herpesvirus 3) is a member of the alphaherpesvirus subfamily of the Herpesviridae (1). The complete DNA sequence of the VZV genome (Dumas strain) was first determined by Davison and Scott in 1986 (2). VZV has a double stranded DNA genome of approximately 125kb with a G+C content of 46%. The genome is composed of a unique long region (UL) flanked by inverted repeat regions namely, the terminal repeat long (TRL) and internal repeat long (IRL) sequences, followed by a unique short (US) region flanked by inverted repeat regions namely, the terminal repeat short (TRS) and internal repeat short (IRS) (Figure 1). The genome encodes 70 unique Open Reading Frames (ORFs), three of them encoded within the IRS region (ORFs 62-64) being inversedly repeated in the TRs region (ORFs 69-71) (1, 2).
VZV is the causative agent of varicella (chickenpox) mostly in children. Subsequently it establishes latency in the sensory ganglia with the potential to reactivate at a later time to cause Herpes Zoster (HZ), also known as shingles. Although varicella is usually self-limiting and resolves within several days, severe complications including pneumonia, encephalitis, congenital infection or even death can occur (1).

Varicella can be effectively controlled and prevented by immunization with live attenuated vaccines. Most live attenuated varicella vaccines were derived from the same parental strain (pOka) of wild type VZV (3, 4). pOka was originally isolated in 1971 from the vesicle fluid of a Japanese child named Oka who had typical chickenpox. The Oka virus was attenuated by sequential passage in cell culture and the resulting Oka strain vaccine (referred as vOka-Biken in this study) was used to vaccinate Japanese children in 1974 (5). Subsequently, GlaxoSmithKline (GSK) Biologicals and Merck Sharp & Dohme (Merck & Co.) have produced their own formulations of the Oka vaccine, known as Varilrix® and Varivax® respectively, which were both derived from the Biken seed stock. Varilrix® was first licensed in some European countries in 1984 whereas Varivax® was approved in the United States in 1995 (4).

A comparative sequence analysis of pOka and vOka using overlapping PCR and Sanger sequencing revealed 42 single nucleotide polymorphisms (SNPs) between the wild type and the attenuated Oka strain (6). More recently, Depledge et al. completed this analysis by comparing pOka with vOka-Biken, vOka-Varilrix from GSK and vOka-Varivax from Merck & Co. (7). The authors generated whole genome sequences (WGS) using deep sequencing technology for one vOka-Biken, two vOka-Varilrix, three vOka-
Varivax and one vOka-Zostavax (related to vOka-Varivax) vaccine preparations. They identified 137 mutated positions present in all vaccine batches. In addition, they showed almost perfect correlation between allele frequencies (AF) in vOka-Biken and vOka-Varilrix whereas vOka-Varivax featured significant differences (7).

In China, varicella vaccine became available in 1998. Currently, the vaccine is produced by four domestic companies, namely Changchun BCHT Biotechnology also known as Baike, Changchun Keygen Biological Products, Changsheng Biotechnology, and Shanghai Institute of Biologic Products. As with Merck & Co. and GSK vaccines, all Chinese varicella vaccines were derived from the Biken seed virus. The vaccines are licensed for the susceptible population which is over 12 months in China (8). All varicella vaccine productions are based on the seed lot system using classical cell culture methods (9). The immunogenicity and safety of the Chinese varicella vaccine have been previously proved in the marketplace (8-11). Adverse event surveillance has not detected fatalities attributable to vaccine in China whereas a few varicella fatal cases in immunocompromised patients have been reported in the USA and Germany and they were both shown to be caused by Oka vaccine strain (12, 13).

The present study aimed at analyzing the Whole Genome Sequence (WGS) of the Chinese varicella vaccine produced by Changchun BCHT biotechnology also known as Baike. In order to better characterize vOka-BK, vOka-BK WGS was compared to the published sequences of VZV wild type strains (Dumas, pOka) and vaccine strains (vOka-Biken, vOka-Varilrix, and vOka-Varivax) in the context of Depledge et al. analysis (7).
RESULTS

Analysis of vOka-BK WGS. vOka-BK WGS was generated with 264 sequencing reads of an average of 717 bp and 94% quality score (94% of the base calls have less than 1% chance of being incorrect). The assembly of an average of 2x fold coverage resulted in a consensus sequence of 125,095 bp (Figure 1A and B). This genome was characterized by a G + C content of 46.1% which is a hallmark of VZV genomes (Figure 1B). As expected, the gene repertoire of vOka-BK was 100% conserved with 70 unique ORFs and 3 ORFs (ORFs 62, 63, 64) inversely duplicated in the TRs region (Figure 1A). The overall genome structure of vOka-BK was also conserved, with the sizes of unique and repeated regions similar to what was observed in the genomes of the other vaccine strains (Figure 1B).

Comparison of vOka-BK WGS with the genome sequence of the parental strain pOka and 3 vaccine strains, vOka-Biken, vOka-Varilrix and vOka-Varivax. vOka-BK WGS (MF898328) was compared with the genome sequences of the parental strain pOka (AB097933) and 3 vaccine strains vOka-Biken (AB097932), vOka-Varilrix (DQ008354), and vOka-Varivax (DQ008355). A total of 168 nucleotide position differences among the 5 Oka strains was identified, including the 137 positions previously identified by Depledge et al. (7) (Supplementary Table 1). Nearly half of the differences (74) were observed in the IRs region, mainly in ORF62. Seventy-two were found in the U1 region. The remaining positions were found in the IRs (18), IRL (3) and US (1) regions. The 18 positions located in the IRs were discarded as similar mutations...
in the repeat motifs have been previously reported in other VZV genomes with unknown consequence on the biology of the virus. The study from Depledge et al. provided AF for sequences of vOka-Biken, vOka-Varilrix and vOka-Varivax genomes (7). Among the 137 SNPs reported by Depledge et al., we chose to exclude 32 sites with AF<10% (7). Furthermore, we excluded 54 sites with low complexity sequence environment such as 4 or more identical nucleotides or dinucleotides repeated multiple times. Finally, we chose to focus on single nucleotide changes. Ten insertions or deletions of one or more nucleotides were detected in non-coding regions and these changes are not likely to have an effect on the vaccine potential. Among the 54 remaining sites, 47 were within ORFs and 28 were non-synonymous substitutions (Figure 1A, Table 1, Supplementary Table 1). ORF 62 and ORF55 featured the most non-synonymous mutations with 9 and 3 respectively.

Comparison of vOka-BK sequence at 54 genomic positions relative to vOka-Biken and vOka-Varilrix and vOka-Varivax. vOka-BK assembly was manually checked at each of the 54 positions of interest. In order to increase confidence in the sequence analysis, these 54 positions were checked on 4 distinct batches, one from 2014 and 3 from 2018. The 3 batches from 2018 yielded the same sequence; not only the main peak was identical but the secondary peak had a similar height (Supplementary Table 2). The 2018 sequence of vOka-BK was compared with the sequence of the other 3 vOka strains. Four profiles were identified relative to the 3 other vaccine strain sequences: 1- twenty-two vOka-BK positions featured the same mutation profile as the other vaccine strains (in grey in all figures and tables); 2- seventeen vOka-BK positions featured the same
mutation profile as vOka-Biken and vOka-Varilrix (in purple); 3- five vOka-BK positions featured the same mutation profile as vOka-Varivax (in pink); 4- ten vOka-BK positions featured a mutation profile different from the other 3 vaccine strains (in brown). Eight of these positions were within ORFs and 6 were non-synonymous changes. They were S945P in the primase subunit of helicase primase ORF6, W87R in glycoprotein (g) N ORF9A, S1167L in the single-stranded DNA-binding protein ORF29, I593V in gB ORF31, A585T in the helicase subunit of the helicase-primase ORF55, and L44P in uracil-DNA glycosylase ORF59.

**Sixteen nucleotide variable positions within ORF62 region.** Thirteen SNPs were found within ORF62 resulting in 9 non-synonymous substitutions (Table 1, Figure 2). Among them, 3 positions (107797, 107599 and 107252) were located within functional sites. The corresponding amino acid changes (L446P, V512A and S628G) were located within the binding site of the transcriptional regulator encoded by ORF63 (Figure 2). In addition, the mutations V512A and S628G were also located in the DNA binding site encoded within ORF62 protein. Furthermore, 2 SNPs (109137, 109200) were found in a putative promoter region of ORF62 gene, located at 4 and 137 nucleotides upstream of the ATG (109133) respectively. Finally, one SNP (105169) was 32 nucleotides downstream of the stop codon (105201) and 22 nucleotides upstream of a putative polyA motif (105147) (Table 1).

Among the 13 SNPs within the ORF62 region, 8 (105544; 105705; 106262; 106933; 107252; 107797; 108111; 108838) had a similar profile in all vOkas including vOka-BK (in grey in Table 1 and Figures 1 and 2). Five of these positions (105544;
105705; 106262; 107252; 108111) were fixed or nearly fixed. Among them, the positions 105705 and 108111 did not result in any amino acid change raising the question why such silent nucleotide changes would be selected. The remaining 3 positions (106933; 107997; 108838) were mixed with wildtype consensus in all vOkas (AF<40% or no distinct SP). Two positions (105310, 105356) were nearly fixed in vOka-Biken and vOka-Varilrix (AF>70) but vOka-Varivax featured a mixed base at these positions. vOka-BK profile was similar to vOka-Biken and vOka-Varilrix with a mutated base and no significant secondary peak (shown in purple in Table1, Figures 1 and 2). In the same category, the position 107599 was mixed in vOka-Biken and vOka-Varilrix whereas it was barely mutated in vOka-Varivax. vOka-BK was mixed at this position (shown in purple in Table1, Figures 1 and 2). One position (106932) was mixed in vOka-Biken and vOka-Varilrix but was almost wildtype (AF=16%) in vOka-Varivax. Similarly, vOka-Bk did not show any mutation at this position as well (shown in pink in Table 1, Figures 1 and 2). Finally, the position 107136 had a mutation profile different in vOka-BK compared to the other vOkas (shown in brown in Table 1, Figures 1 and 2). This position was nearly fixed in vOka-Biken and vOka-Varilrix (AF>95) but it was mixed in vOka-Varivax (AF=53). vOka-BK position was intermediate, mutated with a distinct secondary peak.

Nineteen amino substitutions in 15 proteins other than ORF62. Five substitutions (10900; 58595; 87306; 87815; 115295) were found in 4 glycoproteins (g) including 2 with a surface exposure, namely gB (58595, I593V) and gI (115295, N267S), two
glycoproteins involved in the attachment of the virus (Table 2). The other 3 amino acid substitutions (10900, W87R in gN; 87815, V23A and 87306, S193G in gM) were predicted to be located inside the virion. The rest of the amino substitutions concerned proteins involved in transcription, replication or egress of the virus and were within undetermined functional domains.

Seven positions (488, 560, 85594, 87306, 90535, 97796 and 115295) featured a similar mutation profile in all vOkas including vOka-BK (in grey in Table 2, Figures 1 and 3). One site was particularly intriguing in ORF0. The SNP at position 560 was nearly fixed in all vOkas including vOka-BK and resulted in the extension of the protein for an additional 108 amino acids. The remaining 6 SNPs (488; 85594; 87306; 90535; 97796; 115295) had a similar profile in all vOkas, being mixed with a wild type consensus. Six positions (12779; 31732; 71252; 87815; 97479; 111650) had a similar profile in vOka-Biken, vOka-Varilrix and vOka-BK (in purple in Table 2, Figures 1 and 3). Finally, 6 positions (5745, 10900, 54356; 58595; 97748; 101089) had a unique mutation profile in vOka-BK (in brown in Table 2 and Figure 3). Five of these positions involved proteins that were essential for the biology of the virus. The position 5745, resulting in the amino acid change S945P in the primase ORF6, was nearly fixed in vOka-Biken and vOka-Varilrix whereas it was mixed with a wild type consensus in vOka-Varivax. vOka-BK was mixed at this position with a distinct secondary peak (height at 87). The position 10900, resulting in the amino acid change W87R in gN, was mutated in vOka-BK but it was mixed in vOka-Biken and vOka-Varilrix and wildtype in vOka-Varivax. The position 54356, resulting in the amino acid change
S1167L in the DNA binding protein ORF29, was wild type in vOka-BK but it was mixed in the other vOkas. The position 58595, resulting in the amino acid change I593V in gB, was nearly fixed in vOka-BK whereas it was mixed in the other 3 vOkas. Similarly, the position 97748, resulting in the amino acid change A585T in the helicase ORF55, was nearly fixed in vOka-BK whereas it was mixed in the other 3 vOkas. Finally, the position 101089, resulting in the amino acid change L44P in the uracil DNA-glycosylase ORF59, was mutated in vOka-BK with no distinct secondary peak whereas it was mixed in the other vOkas. Among the 19 substitutions, only one amino acid change, (W87R in ORF9A), was identified as “deleterious” in a PROVEAN analysis (http://provean.jcvi.org/seq_submit.php) meaning that this amino acid substitution was not found in any orthologs (Table 2). The residue 87 is at the very end of the protein and a change from W to R might not be critical for the protein function even though these residues are chemically very different, neutral hydrophobic and basic hydrophilic respectively.

DISCUSSION

VZV vaccine is a live attenuated vaccine. Attenuation was obtained after multiple passages in cultured cells (5). Most of the vaccines were derived from clade 2 strain pOka and were passaged at least 30 times in various cell types (4). Comparisons between parental and vaccine genomes identified potential sites that could be associated with attenuation (6, 7). Additional studies on the clade 3 strain Ellen, which became highly attenuated after at least 90 passages in cultured cells, showed that position 560
within ORF0 and positions 106262 and 107252 with ORF62 were likely to be associated with attenuation (14). Attenuation was confirmed using SCID-hu mouse model (15). Infection with vOka and Ellen leads to minimal replication compared to infection with pOka. The mutation at position 560 results in a longer ORF0, 221 amino acids instead of 129. ORF0, also known as ORF S/L, encodes a cytoplasmic protein that might affect adhesion molecules in infected cells (16). More recently, Zhang et al., using the Bacterial Artificial Chromosome (BAC) technology, showed that the deletion of the entire ORF resulted in a slow-growth phenotype in infected cells (17). Finally, Kaufer et al. demonstrated that ORF0 is essential for VZV replication and could be considered as functional homolog of HSV-1 UL56 (18). Position 107252, corresponding to residue 628 in ORF62 is within the DNA binding site (467-640) whereas position 106262 (residue 958) has not been associated to any functional domain of ORF62 (19). Recently, a comparison between WGS from pOka and WGS from 7 distinct vOka preparations showed that positions 560, 106262 and 107252 were fixed or nearly fixed (7). The current study confirmed that these positions are mutated in vOka-BK. As these 3 mutations are shared between two attenuated viruses, vOka and Ellen strain, this suggests that these positions might be necessary for the attenuation process. A definite demonstration would be to include these mutations in a wild type virus and assess whether the recombinant virus is attenuated. Three additional positions within ORF62 are fixed or nearly fixed, namely 105544, 105705, 108111. Cohrs et al. showed that transactivation of VZV promoters with ORF62 derived from vOka or pOka were comparable suggesting that vOka ORF62 was not sufficient to induce attenuation.
Furthermore, Zerboni et al. showed that attenuation is likely to involve different regions of the genome (21). The comparison of vOka genomes with pOka identified other positions in vOkas that have the same mutation profile, basically mixed with a wildtype consensus. Nine of these nucleotide substitutions resulted in amino acid changes but none were identified as deleterious in a PROVEAN analysis suggesting that these amino acid substitutions could be found in other strains. It would be however worth assessing whether a fixation of these positions would have an effect on the attenuated phenotype.

Varilrix® and Varivax® are 2 of the most used VZV vaccines in the world (4). Even though they are both derived from pOka, a recent study showed that vaccine preparations from Varilrix® and Varivax® are not genetically identical. Indeed, a correlation of 0.69 was found between vOka-Biken and vOka-Varivax AFs (7). In contrast, Depledge et al reported an almost perfect correlation between vOka-Biken and vOka-Varilrix AFs ($R^2=0.98$). Varilrix® and Varivax® have been established after multiple passages in various cell lines and the discrepancy might be due to passage differences as reported by Depledge et al. (7). The present study identifies 22 positions that might be related to passage history. Several studies compared Varilrix® and Varivax® in terms of vaccine efficacy and adverse reaction. Higher seroconversion rates have been reported for Varivax® (22, 23). Furthermore, higher risk of varicella breakthrough was reported in patients vaccinated with Varilrix® (24). The present study shows that vOka-BK has more sites with a profile like vOka-Biken and vOka-Varilrix. That might suggest that vOka-BK might have a similar vaccine potential compared to
Reports concerning Baike vaccine specifically are very scarce. Wang et al. reported a vaccine efficacy of 91% (Confidence Interval, CI: 67%-93%) for Baike vaccine which is comparable to the average vaccine efficacy (95% CI: 78%-84%) computed in a meta-analysis conducted by Marin et al. based on 42 original studies (9, 25). As far as we know, the only publication concerning specifically the Baike vaccine was published in the Chinese Journal of Immunology by Dong et al (26). Briefly, this study concerned 38,883 children from the Binhai New Area in the city of Tianjin, bordering Beijing in the Southeast, aged from 12 months to 12 years old, which were monitored from 42 days to 2 years after vaccination with the Baike vaccine. Dong et al. reported 134 breakthrough cases (0.34%). This incidence rate can be compared to the average incidence of 0.85% (CI: 0.53%-1.37%) computed recently in a meta-analysis of 27 studies by Zhu et al. (27). Thus, reports from Wang et al. as well as Dong et al. suggest that Baike vaccine is very efficacious. However, more data is needed especially for 2 dose vaccination as the reports mentioned concerned only 1 dose vaccination. In this context, the 10 SNPs that have a profile different in vOka-BK compared to the other Oka vaccine might be involved in a better vaccine efficacy. The present study shows some batch variability at least between year of production. The lots from 2018 seem comparable. In order to have a better sense of vOka-BK vaccine efficacy, it would be necessary to do a side by side comparison with the other vOkas.

Currently, many countries have introduced varicella vaccine in universal routine vaccination (URV) program and administered 2 doses of vaccine to control and prevent...
chickenpox (28). WHO recommends varicella vaccine as URV in the countries and regions where there is a heavy burden of disease. At this time, varicella vaccine is only available in the private sector in China. VZV vaccine has been proved safe and well tolerated (29, 30). Unfortunately, several cases reported that varicella live attenuated vaccine administrated in immunocompromised patients could occasionally cause severe adverse events and even death (12, 13). Though very rare, these lethal cases highlight the challenge of a safe administration of live VZV vaccine in immunocompromised patients. It is worth noting that no death has been reported in immunocompromised patients from China. This is likely due to a gap of surveillance and this highlights the need for better monitoring of varicella outbreaks in China.

In summary, the comparison of WGS from vOka-BK used in the varicella vaccine preparation manufactured by the Chinese company Changchun BCHT Biotechnology, also known as Baike, with 3 other vOkas WGS identified 3 types of SNPs. 1- Nucleotide positions that featured a similar mutation profile, either mutated or wildtype in all vOkas and are likely to be core vaccine positions. A change at these positions might be detrimental for the vaccine potential. 2- SNPs that featured a different profile in vOka-Varilrix and vOka-Varivax and could be associated with differences in vaccine efficacy. 3- And finally, positions that have a unique profile in vOka-BK and could affect the vaccine efficacy of the Chinese vaccine.

MATERIAL AND METHODS

Vaccine lots and DNA extraction. vOka-BK was produced by Changchun BCHT
Biotechnology, (Changchun, China) from vOka-Biken seed stock in MRC5 ATCC cells (5). The number of passages in MRC5 cells is unknown but likely to be less than 12 to limit the number of pOka passages to 38 as required by WHO (31). Four vaccine lots of vOka-BK have been used. One lot produced in 2014 (201402015-1) and 3 lots produced in 2018 (201802011-2, 201805038-1, and 201805039-1). Whole genomic DNA with the QIAamp DNA Mini Kit (Qiagen, Düsseldorf, Germany) following the manufacturer’s instructions.

**PCR amplification.** Forty-two sequence-specific primers were designed using the Sequencher 5.0 software (GeneCodes, Ann Arbor, MI) based on the sequence of the reference strain Dumas (GenBank accession no. X04370) (Figure 1). Overlapping regions of approximately 200 bases were designed to cover the entire genomic sequence of VZV. Overlapping long PCR fragments (4.0–6.5 kb) were amplified with the high-fidelity enzyme mixture of the Platinum Taq DNA Polymerase High Fidelity kit following manufacturer’s instructions (Invitrogen, Carlsbad, CA). The thermal cycler program consisted of an initial hot-start PCR step of 94°C for 2 min followed by 40 cycles of amplification (94°C for 15 s, 50°C for 30 s, and 68°C for 6.5 min) and a final elongation step at 68°C for 5 min. PCR products were then purified using the QIAquick PCR purification and by the Gel Extraction kit (Qiagen, Düsseldorf, Germany) when necessary. In addition, 24 amplicons were designed to specifically cover the 54 genomic positions of interest. PCR fragments (0.8–2.0 kb) were amplified with the high-fidelity enzyme mixture of the Platinum Taq DNA Polymerase High Fidelity kit following manufacturer’s instructions (Invitrogen, Carlsbad, CA). The thermal cycler program
was as follows: an initial hot-start PCR step of 94°C for 2 min was followed by 40 cycles of amplification (94°C for 15 s, 50°C for 30 s, and 68°C for 2 min) and a final elongation step at 68°C for 5 min. The sequences of primers used for amplification as well as sequencing are available upon request.

**DNA sequencing.** Direct sequencing of purified PCR products and plasmid DNA was performed with the ABI PRISM Dye Terminator reaction kit (Perkin-Elmer, Norwalk, USA) and an ABI PRISM 3100 genetic analyzer (Applied Biosystems, Hitachi, Japan) according to the manufacturer’s instructions. Sequencing reads were assembled with the Sequencher software version 5.0 (GeneCodes, Ann Arbor, MI). The positions with ambiguous nucleotides were identified using the default value of 15% for detection of secondary peaks in Sequencher.

**Cloning of PCR products.** Direct sequencing of the PCR products derived from regions with highly complex secondary structure (flanking regions between IR$_L$ and IR$_S$ regions) as well as the genomic termini was complemented with sub-cloning of amplicons followed by sequencing of plasmid clones. Sub-cloning was also performed when direct sequencing did not generate information of sufficient quality or when a particular SNP could not be reliably confirmed. PCR amplicons were individually cloned using the TOPO TA cloning kit (Invitrogen, San Diego, USA). Plasmid DNAs were purified from 7 bacterial clones with a QIAprep spin kit (Qiagen, Düsseldorf, Germany). DNA sequences of the cloned inserts were determined using vector-specific sequencing primers.

**Sequence analysis.** vOka-BK WGS was compared with the following 5 genomic
sequences from GenBank database (Dumas X04370, pOka AB097933, vOka-Biken AB097932, vOka-Varilrix DQ008354, and vOka-Varivax DQ008355). All described nucleotide and protein positions in this study are relative to the reference Dumas strain, X04370. Sequences were aligned with MAFFT (32), SNPs were listed in MEGA 6 (33) and sequence alignments were analyzed with BioEdit version 7.0 (Tom Hall, North Carolina State University, Raleigh, NC). Genomic maps were generated with Artemis 16.0.0 (34). Average AFs of vOka-Varilrix and vOka-Varivax were computed based on the data reported by Depledge et al. from 2 GSK (Varilrix®) and 4 Merck & Co. (3 Varivax®, 1 Zostavax®) vaccine preparations (7).

**Accession number:** Oka-BK WGS is available in GenBank under the accession number MF898328.

**Acknowledgements**

This work was supported by the Key Technologies R&D Program of the National Ministry of Science (2018ZX10713002 and 2018ZX10713001-003), and the National Key Research and Development Program (2016YFC1200905).

Declarations of interest: none

**REFERENCES**

1. Arvin AM, Gilden D. 2013. Varicella-Zoster Viruses, p. 2015-2057. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed.), *Fields Virology*, 5th ed, vol. 6. Lippincott, Williams & Wilkins, Philadelphia.
2. Davison AJ, Scott JE. 1986. The complete DNA sequence of varicella-zoster virus. The Journal
of general virology 67 (Pt 9):1759-1816.

3. Kim JI, Jung GS, Kim YY, Ji GY, Kim HS, Wang WD, Park HS, Park SY, Kim GH, Kwon SN, Lee KM, Ahn JH, Yoon Y, Lee CH. 2011. Sequencing and characterization of Varicella-zoster virus vaccine strain SuduVax. Virol J 8:547.

4. Tilieux SL, Halsey WS, Thomas ES, Voycik JJ, Sathe GM, Vassilev V. 2008. Complete DNA sequences of two oka strain varicella-zoster virus genomes. J Virol 82:11023-11044.

5. Takahashi M, Otsuka T, Okuno Y, Asano Y, Yazaki T. 1974. Live vaccine used to prevent the spread of varicella in children in hospital. Lancet (London, England) 2:1288-1290.

6. Gomi Y, Sunamachi H, Mori Y, Nagaike K, Takahashi M, Yamanishi K. 2002. Comparison of the complete DNA sequences of the Oka varicella vaccine and its parental virus. J Virol 76:11447-11459.

7. Depledge DP, Yamanishi K, Gomi Y, Gershon AA, Breuer J, Frueh K. 2016. Deep Sequencing of Distinct Preparations of the Live Attenuated Varicella-Zoster Virus Vaccine Reveals a Conserved Core of Attenuating Single-Nucleotide Polymorphisms. Journal of Virology 90:8698-8704.

8. Lu X, Zhou W, Gu K, Zhu F. 2008. Evaluation on safety and immunogenicity of Changshengkeji frozen-dried live attenuated varicella vaccine. South China J Prev Med 207-211.

9. Wang Z, Yang H, Li K, Zhang A, Feng Z, Seward JF, Bialek SR, Wang C. 2013. Single-dose varicella vaccine effectiveness in school settings in China. Vaccine 31:3834-3838.

10. Chen E, Jiay Z, Li Q. 2009. Safety and immunogenicity of lyophilized live attenuated domestic varicella vaccine. Zhongguo yi miao he mian yi 15:435-437.

11. Fu C, Wang M, Liang J, Xu J, Wang C, Bialek S. 2010. The effectiveness of varicella vaccine in China. Pediatr Infect Dis J 29:690-693.

12. Jessica L, Subhadra S, Jones JF, Cynthia S, Debra B, D Scott S, Bialek SR, Mona M. 2014. Fatal varicella due to the vaccine-strain varicella-zoster virus. Human Vaccines & Immunotherapeutics 10:146-149.

13. Schrauder A, Henkegendo C, Seidemann K, Sasse M, Cario G, Moericke A, Scharpp M, Heim A, Wessel A. 2007. Varicella vaccination in a child with acute lymphoblastic leukaemia. Lancet (London, England) 369:1232-1232.

14. Peters GA, Tyler SD, Carpenter JE, Jackson W, Mori Y, Arvin AM, Grose C. 2012. The attenuated genotype of varicella-zoster virus includes an ORF0 transitional stop codon mutation. J Virol. 86:10695-10703. doi: 10610.1128/JVI.01067-10612. Epub 12012 Jul 02620.

15. Moffat JF, Zerboni L, Kinchington PR, Grose C, Kaneshima H, Arvin AM. 1998. Attenuation of the vaccine Oka strain of varicella-zoster virus and role of glycoprotein C in alphaherpesvirus virulence demonstrated in the SCID-hu mouse. J Virol. 72:965-974.

16. Kemble GW, Annunziato P, Lungo O, Winter RE, Cha TA, Silverstein SJ, Spaete RR. 2000. Open reading frame S/L of varicella-zoster virus encodes a cytoplasmic protein expressed in infected cells. J Virol. 74:11311-11321.

17. Zhang Z, Rowe J, Wang W, Sommer M, Arvin A, Moffat J, Zhu H. 2007. Genetic analysis of varicella-zoster virus ORF0 to ORF4 by use of a novel luciferase bacterial artificial chromosome system. J Virol. 81:9024-9033. doi: 9010.1128/JVI.02666-02606. Epub 02007 Jun 02620.

18. Kaufer BB, Smekal B, Osterrieder N. 2010. The varicella-zoster virus ORFS/L (ORF0) gene
is required for efficient viral replication and contains an element involved in DNA cleavage. J Virol. 84:11661-11669. doi: 11610.11128/JVI.00878-11610. Epub 12010 Sep 11615.

19. Tyler JK, Everett RD. 1994. The DNA binding domains of the varicella-zoster virus gene 62 and herpes simplex virus type 1 ICP4 transactivator proteins heterodimerize and bind to DNA. Nucleic Acids Res. 22:711-721.

20. Cohrs RJ, Gilden DH, Gomi Y, Yamanishi K, Cohen JI. 2006. Comparison of virus transcription during lytic infection of the Oka parental and vaccine strains of Varicella-Zoster virus. J Virol. 80:2076-2082. doi: 11610.1128/JVI.2080.2075-2082.2006.

21. Zerboni L, Hinchliffe S, Sommer MH, Ito H, Besser J, Stamatis S, Cheng J, Distefano D, Kraiouchkine N, Shaw A, Arvin AM. 2005. Analysis of varicella zoster virus attenuation by evaluation of chimeric parent Oka/vaccine Oka recombinant viruses in skin xenografts in the SCIDhu mouse model. Virology. 332:337-346. doi: 310.1016/j.virol.2004.1010.1047.

22. Blatter MM, Klein NP, Shepard JS, Leonardi M, Shapiro S, Schear M, Mufson MA, Martin JM, Varman M, Grogg S, London A, Cambron P, Douha M, Nicholson O, da Costa C, Innis BL. 2012. Immunogenicity and safety of two tetravalent (measles, mumps, rubella, varicella) vaccines coadministered with hepatitis a and pneumococcal conjugate vaccines to children twelve to fourteen months of age. Pediatr Infect Dis J. 31:e133-140. doi: 110.1097/INF.1090b1013e318259fc318258a.

23. Lau YL, Vessey SJ, Chan IS, Lee TL, Huang LM, Lee CY, Lin TY, Lee BW, Kwan K, Kasim SM, Chan CY, Kaplan KM, Distefano DJ, Harmon AL, Golie A, Hartzel J, Xu J, Li S, Matthews H, Sadoff JC, Shaw A. 2002. A comparison of safety, tolerability and immunogenicity of Oka/Meck varicella vaccine and VARILRIX in healthy children. Vaccine. 20:2942-2949.

24. Spackova M, Wiese-Posselt M, Dehnert M, Matysiak-Klose D, Heininger U, Siedler A. 2010. Comparative varicella vaccine effectiveness during outbreaks in day-care centres. Vaccine. 28:686-691. doi: 610.1016/j.vaccine.2009.1010.1086. Epub 2009 Oct 1027.

25. Lynch JM, Kenyon TK, Grose C, Hay J, Ruyechan WT. 2002. Physical and functional interaction between the varicella zoster virus IE63 and IE62 proteins. Virology. 302:71-82.

26. Dong XJ, Liu YP, Wang LJ, Wang WM, Yang HJ, Li YC, Wang W, Gang HaY, Zhuang ZR. 2014. Evaluation of immune effect of Tianjin Binhai New Area of varicella vaccine. Chinese Journal of Immunology:1114-1116,1120.

27. Zhu S, Zeng F, Xia L, He H, Zhang J. 2018. Incidence rate of breakthrough varicella observed in healthy children after 1 or 2 doses of varicella vaccine: Results from a meta-analysis. American journal of infection control 46:e1-e7.

28. Wutzler P, Bonanni P, Burgess M, Gershon A, Safadi MA, Casabona G. 2017. Varicella vaccination - the global experience. Expert Rev Vaccines. 16:833-843. doi: 810.1080/14760584.14762017.11343669. Epub 14762017 Jul 14760513.

29. Pileggi GS, Ferriani VPL. 2010. Safety and immunogenicity of varicella vaccine in patients with juvenile rheumatic diseases receiving methotrexate and corticosteroids. Arthritis Care & Research 62:1034-1039.

30. Zhang J, Xie F, Delzell E, Chen L, Winthrop KL, Lewis JD, Saag KG, Baddley JW, Curtis JR. 2012. Association between vaccination for herpes zoster and risk of herpes zoster infection among older patients with selected immune-mediated diseases. Jama 308:43-49.

31. D’Hondt E, Berge E, Colinet G, Duchene M, Peertmans J. 1985. Production and quality
control of the Oka-strain live varicella vaccine. Postgrad Med J 61:53-56.

32. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research 30:3059-3066.

33. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729.

34. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16:944-945.

35. Ruyechan WT. 2010. Roles of cellular transcription factors in VZV replication. Curr Top Microbiol Immunol 342:43-65.10.1007/1082_2010_1042.

FIGURE LEGENDS

Figure 1: Comparison of VZV genomes. A. Schematic representation of VZV genome structure and gene repertoire, as well as the sequence strategy used in this report and the nucleotide variability between parental and vaccine strains. Genomic regions are represented as color-coded boxes, in blue for unique regions (dark for U1 and light for Us), in green for repeat long (TR1 and IR1), in grey for internal repeats (IR1, 2, 3, 4a, 4b and 5) and in orange for repeat short (IRs and TRs). ORFs are shown as grey arrows, rightward ORFs on the top and leftward ORFs on the bottom. PCR amplicons are shown as thick lines. Fifty-four SNPs in vOka strains compared to pOka are shown as arrowheads for non-synonymous substitutions and as “X” for synonymous substitutions. The 54 SNPs are color-coded based on the 4 vOka-BK categories established in this analysis, in grey vOka-BK positions with the same profile as the 3 other vaccine strains, in purple vOka-BK positions with similar profile than vOka-Biken and vOka-Varilrix, in pink vOka-BK positions with similar profile than vOka-Varivax and in brown vOka-BK positions with a unique profile. B. Genomic components of the Oka genomes analyzed in this study. Elements outside of the range of the average size are shown in
Figure 2: Thirteen SNPs within ORF62. Each SNP is indicated by its genomic position and represented by a box featuring a representative chromatogram for vOka-BK sequencing data as well as the alignment between pOka, vOka-Biken, vOka-Varilrix, vOka-Varivax and vOka-BK. The nucleotide of interest is shown in bold in the alignment. Allele frequencies (F) and secondary peak (SP) are shown for the nucleotide position. A secondary peak below detection level is indicated as ND. Amino acid residues (in red) and positions within ORF62 are indicated. Except for the box representing the positions 106932 and 106933, each box is outlined with a thick line in color based on the 4 vOka-BK categories established in this analysis: in grey vOka-BK positions with the same profile as the 3 other vaccine strains, in purple vOka-BK positions with similar profile than vOka-Biken and vOka-Varilrix, in pink vOka-BK positions with similar profile than vOka-Varivax and in brown vOka-BK positions with a unique profile. For the box corresponding to the genomic positions 106932 and 106933, the column corresponding to allele frequency and secondary peak data is boxed. Positions corresponding to non-synonymous changes are shown as a thick line on the genome diagram. The bottom of the figure features a functional map drawn on scale based on the review from Ruyechan, 2010 (35). NLS: Nuclear Localization Signal; IE: Immediate Early; S: region rich in serine; pS: Phospho-serine.

Figure 3. Nineteen non-synonymous substitutions (excluding ORF62). The layout of
the figure is according to Figure 2. The bottom of the figure features a genomic map in order to better locate the non-synonymous substitutions within the genome and its ORFs. Genomic regions are represented as color-coded boxes, in blue for unique regions (dark for UL and light for US), in green for repeat long (TRL and IR1), in grey for internal repeats (IR1, 2, 3, 4a, 4b and 5) and in orange for repeat short (IRs and TRs). ORFs are shown as grey arrows, rightward ORFs on the top and leftward ORFs on the bottom.
| Position | Change Type | Protein Function | AF (%) | SD (%) |
|----------|-------------|------------------|--------|--------|
| 107599  | neutral→non-p | transcriptional regulator ICP4 | 57.7  | 17.9   |
| 107797  | neutral→non-p | transcriptional regulator ICP4 | 24.1  | 29.7   |
| 111650  | neutral→non-p | virion prot. US10 | 73.2  | 29.3   |
| 112331  | neutral→non-p | transcriptional regulator ICP4 | 57.7  | 17.4   |
| 84091   | neutral→non-p | tegument | 28  | 26.7   |
| 89734   | neutral→non-p | helicase | 22.1  | 26.3   |
| 26125   | neutral→non-p | helicase | 73.2  | 75.8   |
| 85594   | neutral→non-p | helicase-primase subunit | 25.2  | 35.4   |
| 90535   | neutral→non-p | helicase-primase subunit | 25.2  | 35.4   |
| 84084   | neutral→non-p | helicase-primase | 31.4  | 29.6   |
| 89734   | neutral→non-p | helicase-primase | 41.1  | 40.6   |
| 26125   | neutral→non-p | helicase-primase | 73.2  | 75.8   |
| 85594   | neutral→non-p | helicase-primase | 25.2  | 35.4   |
| 90535   | neutral→non-p | helicase-primase | 25.2  | 35.4   |

**Table 1:** List of 54 SNPs in vOka-Biken, v-Oka-Varilrix, vOka-Varivax and vOka-BK compared to pOka
1- The 54 SNPs are sorted based on the 4 vOka-BK categories established in this analysis:
in grey vOka-BK positions with the same profile as the 3 other vaccine strains, in purple
vOka-BK positions with similar profile than vOka-Biken and vOka-Varilrix, in pink vOka-
BK positions with similar profile than vOka-Varivax and in brown vOka-BK positions with
a unique profile. The positions are also sorted based on the allele frequency of vOka-Biken.
Three positions 560, 106262 and 107252 are shared with the clade 3 Ellen strain, known
to be attenuated, and are indicated with a star * (Peters et al. 2012).
2-SNPs within ORFs are indicated by the ORF name (number) whereas SNPs outside ORF
are indicated as NCR (Non-Coding Region).
3- abbreviations: prot.: protein; g: glycoprotein; ss: single stranded; S/T PK:
Serine/Threonine Protein Kinase.
4- pOka nucleotide is shown as well as the consensus nucleotide for the 3 vOkas. Mixed
nucleotides are shown unless allele frequency is < 5% (wild type) or >95% (mutated).
5- Average allele frequencies (AF) are represented as heat map, from green for wild type
(AF=0%) to red for fixed mutation (AF=100%) through orange (AF=50%). Average AF
values were computed using raw data from Depledge et al, 2 vOka-Varilrix vaccine batches
on one hand, 3 vOka-Varivax batches and 1 Zostavax batch on the other hand (Depledge
et al., 2016). SD: Standard Deviation.
6- Positions with SD > 10: AF values for vOka-Varilrix, 54354 (0%; 21.74%), 54356 (0%;
31.82%), AF values for vOka-Varivax, 86478 (17.24%; 30.40%; 36.15%; 54.17%), 54354
(0%; 25.64%; 27.55%; 0%), 106932 (36.16%; 0.96%; 1.84%; 25.11%), 54356 (29.41%;
42.24%; 38%; 66.67%). Based on Depledge et al., 2016.
7- The base corresponding to the main peak as well as to a potential secondary peak is
shown for vOka-BK.

SNPs within ORFs are indicated by their amino position. Non-synonymous changes are indicated as well as the type of change in terms of polarity (p: polar) and hydrophobicity (hl: hydrophilic; hb: hydrophobic).
| Genomic position | ORF | Annotation | Main function | Location | Essential? | INP | Allele frequency | SNP Allele vs. | Description of structure/domain including the amino acid residue of interest | PROVEAN analysis |
|------------------|-----|------------|---------------|----------|-----------|-----|-----------------|---------------|--------------------------------------------------------------------------|------------------|
| 54805            | 26  | gB         | attachment and fusion | envelope | yes       | 1   | 0.6             | 0.4           | virology, no known domain                                                  | neutral          |
| 112345           | 47  | gI         | attachment | envelope | impaired   | 1   | 0.7             | 0.9           | virology, no known domain                                                  | neutral          |
| 87063            | 52  | gH         | assembly and egress | impaired | 2         | V35A | 1.1             | 2             | interconversion, no known domain                                           | neutral          |
| 87039            | 54  | gH         | assembly and egress | impaired | 2         | S100G | 18.4            | 25.5          | wild type                                                                | neutral          |
| 19000            | NA  | gM         | maturation of gH | yes      | 1         | W417K | 46.7            | 49.8          | neutral, no known domain                                                   | neutral          |
| 15779            | 20  | transactivating protein VP16 | tegument | no       | 1         | A207V | 20.6            | 20.6          | wild type, distinct SP                                                     | neutral          |
| 488              | 1A  | gN         | maturation of gM | yes      | 1         | *L90R | 66.2            | 66.2          | neutral, no known domain                                                   | neutral          |
| 560              | 1B  | gN         | maturation of gM | yes      | 1         | S945P | 94.9            | 93.4          | neutral, no known domain                                                   | neutral          |
| 5741             | 6   | helicase-primase subunit | primase | yes      | 1         | SHGSP | 94.0            | 85.6          | no known domain                                                            | neutral          |
| 86530            | 52  | helicase-primase subunit | helicase-primase complex | yes | 1         | IS30V | 25.2            | 36.6          | wild type, no known domain                                                  | neutral          |
| 67678            | 35  | helicase-primase subunit | helicase | yes | 1         | V859A | 16.3            | 17.0          | wild type, no known domain                                                  | neutral          |
| 67706            | 55  | helicase-primase subunit | helicase | yes | 3         | A50ST | 47.7            | 46.8          | wild type, no known domain                                                  | neutral          |
| 95594            | 48  | DNAK       | nucleosome assembly and egress | replication | yes | 1         | T310A | 33.9            | 36.5          | wild type, no known domain                                                  | neutral          |
| 54586            | 20  | ssDNA-binding protein | replication | yes | 1         | S1167L | 30.5            | 31.7          | wild type, no known domain                                                  | neutral          |
| 151839           | 59  | US10       | no | yes | 1         | L44P  | 63.2            | 51.1          | no known domain                                                            | neutral          |
| 111630           | 64  | UL56       | no | no | 1         | Q29R  | 73.2            | 75.6          | no known domain                                                            | neutral          |
| 72572            | 39  | UL20       | envelope | yes | 1         | M227T | 76.6            | 80.2          | no known domain                                                            | neutral          |
| 21172            | 21  | UL37       | envelope | yes | 1         | T32S  | 20.3            | 20.5          | wild type, distinct SP                                                     | neutral          |

Table 2: list of 19 non-synonymous substitutions (excluding ORF62)
1- The proteins are listed based on their role in the biology of the virus, attachment, entry, transcription, replication and egress.

2- Annotation, main function, location in the virion and previous results on the role of the protein in the biology of the virus according to Arvin and Gilden, 2013. g: glycoprotein; ss: single stranded.

3- Average allele frequencies are represented as heat map, from green for wild type (AF=0%) to red for fixed mutation (AF=100%) through orange (AF=50%). Average AF values were computed using raw data from Depledge et al (Depledge et al., 2016).

4- The status of vOka-BK sequence is indicated according to the 4 vOka-BK categories established in this analysis: in grey vOka-BK positions with the same profile as the 3 other vaccine strains, in purple vOka-BK positions with similar profile than vOka-Biken and vOka-Varilrix and in brown vOka-BK positions with a unique profile.
Figure 1
Figure 2
