

Qiuwei Pan ORCID iD: 0000-0001-9982-6184

Mohamad Hakim ORCID iD: 0000-0001-8341-461X

## **Phylogenetic and immunoinformatic analysis of VP4, VP7, and NSP4 genes of rotavirus strains circulating in children with acute gastroenteritis in Indonesia**

**Hera Nirwati<sup>1,‡</sup>, Celeste M. Donato<sup>2,3,‡</sup>, Aqsa Ikram<sup>4,5,‡</sup>, Abu T. Aman<sup>1</sup>, Tri Wibawa<sup>1</sup>,  
Carl D. Kirkwood<sup>3</sup>, Yati Soenarto<sup>6</sup>, Qiuwei Pan<sup>4</sup>, Mohamad S. Hakim<sup>1,4,§</sup>**

<sup>1</sup> Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia.

<sup>2</sup> Biomedicine Discovery Institute and Department of Microbiology, Monash University, Melbourne, Victoria, Australia

<sup>3</sup> Enteric Diseases Group, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.

<sup>4</sup> Postgraduate School Molecular Medicine, Erasmus MC-University Medical Center, Rotterdam, the Netherlands.

<sup>5</sup> Atta-Ur-Rahman School of Applied Biosciences, National University of Science and Technology, Islamabad, Pakistan

This is the author manuscript accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/jmv.25527](https://doi.org/10.1002/jmv.25527).

This article is protected by copyright. All rights reserved.

<sup>6</sup> Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada and Dr. Sardjito Hospital, Yogyakarta, Indonesia.

‡ These authors contribute equally to this manuscript

**§ Corresponding author:**

**Mohamad S. Hakim, PhD.**

Department of Microbiology, Faculty of Medicine, Public Health and Nursing

Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

Phone/Fax: +62 274 580297; Email: m.s.hakim@ugm.ac.id.

**Short title:** Phylogenetic and immunoinformatic analysis of Indonesian rotavirus

**ABSTRACT**

Rotavirus is a major cause of diarrhea in Indonesian children. However, rotavirus vaccines have not been introduced in the national immunisation program of Indonesia. Understanding the genetic diversity and conserved antigenic regions of circulating strains are therefore essential to assess potential efficacy of rotavirus vaccines. We collected fecal samples from hospitalised children less than 5 years of age with acute diarrhea. Rotavirus genotyping was performed by reverse transcriptase polymerase chain reaction (RT-PCR), followed by sequencing of the VP4, VP7, and NSP4 genes of representative strains. Phylogenetic analysis was performed to investigate their relationship with globally circulating strains. Conservational analysis, immunoinformatics, and epitope mapping in comparison to vaccine strains were also performed. The sequence analyses showed that differences of multiple amino acid

residues existed between the VP4, VP7, and NSP4 antigenic regions of the vaccine strains and the Indonesian isolates. However, many predicted conserved epitopes with higher antigenicity were observed in the vaccine and Indonesian strains, conferring the importance of these epitopes. The identified epitopes showed a higher potential of rotavirus vaccine to be employed in Indonesia. It could also be helpful to inform the design of a peptide vaccine based on the conserved regions and epitopes in the viral proteins.

**Keywords:** epitopes, Indonesia, genotyping, phylogenetic analysis, rotavirus.

## 1. INTRODUCTION

Infectious diseases, such as pneumonia, diarrhea, and malaria, are the leading cause of worldwide morbidity and mortality in children less than 5 years of age.<sup>1,2</sup> Recent estimates demonstrated that rotavirus is the most frequent cause of severe diarrhea episodes resulting in death among children less than 5 years of age, followed by *Enteropathogenic Escherichia coli* (EPEC) and calicivirus.<sup>3</sup> It was estimated in 2013 that rotavirus infection caused approximately 215,000 deaths in children less than 5 years of age.<sup>4</sup>

The rotavirus genome consists of 11 double-stranded RNA segments which encode 6 structural proteins (viral proteins (VPs)) and 6 non-structural proteins (NSPs). The outer layer is comprised of two neutralising antigens, VP7 and VP4 proteins. These two genes are the basis for the binary classification of rotavirus into G- and P-genotype, respectively.<sup>5</sup> Since these two antigens are encoded by two separate RNA segments, they are able to undergo reassortment between different human and/or animal rotavirus strains, creating new G-P genotype combinations and ultimately leading to the

generation of rotavirus diversity.<sup>6</sup> In fact, novel G- and P-genotype combinations as well as reassortants of human and animal strains have been continuously reported.<sup>7-11</sup>

Indonesia is among the South-East Asian countries with a high burden of rotavirus diarrhea-related deaths.<sup>12</sup> As a member of the Asian Rotavirus Surveillance Network, a national surveillance program was established to monitor the circulation of rotavirus strains in the country.<sup>13-15</sup> The national surveillance conducted in 2006, and between 2009 and 2015 showed that rotavirus was responsible for approximately 47.5%-52.0% of hospital admissions due to acute diarrhea in children under 5 years of age.<sup>13,16</sup> Genotyping studies showed that G1P[8], G1P[6] and G2P[4] strains were the most common cause of rotavirus diarrhea in Indonesia<sup>13,14</sup>, in agreement with findings of worldwide studies.<sup>17</sup> Interestingly, changes of dominant strains circulating in Indonesia were observed by year<sup>13</sup>, including the predominant of G3P[8] since 2013.<sup>16</sup> These findings support the necessity of continuous surveillance in the country.

Currently, there are two commercially available live-attenuated rotavirus vaccines, Rotarix and RotaTeq, that differ in their antigenic specificity. Rotarix (GlaxoSmithKline Biologicals) is comprised of a monovalent G1P[8] strain, derived from the human rotavirus strain RIX 4414.<sup>18</sup> RotaTeq (Merck) is a human-bovine reassortant vaccine, containing five reassortant strains derived from human (G1, G2, G3, and G4) and bovine WC3 (P[8]) strains.<sup>18</sup> Both vaccines provide effective protection against severe rotavirus diarrhea in many countries.<sup>19,20</sup> However, some reports indicated a reduced efficacy of rotavirus vaccines in developing, low-income countries in Asia and sub-Saharan Africa.<sup>21,22</sup> The marked variability of antigenic rotavirus proteins has posed a serious obstacle in the effectiveness of rotavirus vaccines.<sup>23</sup>

Although rotavirus detection and genotyping were regularly conducted in Indonesia<sup>13,14,16</sup>, no further sequence analysis of Indonesian rotavirus strains has been performed. Consequently, the genetic evolution of rotavirus strains over time in Indonesia is unknown. Therefore, the objective of this study is to better understand the genetic variations of rotavirus strains circulating in Indonesia in 2006 and 2009. During the period, mass vaccination against rotavirus was not widely implemented. Rotarix and RotaTeq vaccines have been commercially available in Indonesia since 2011. However, they have not been included yet in the Indonesia's National Immunisation Program (NIP).<sup>16</sup> Therefore, the diversity of rotavirus genes observed in this study was mostly due to the natural evolution of rotavirus in Indonesian pediatric population.

In the present study, we compared the VP4, VP7, and NSP4 genes of human rotavirus strains circulating in Indonesia with global strains and performed phylogenetic analysis. In addition, immunoinformatic analysis was performed to investigate the potential antigenic differences between Indonesian and vaccine strains (Rotarix and RotaTeq). Conservational analysis and epitope mapping were employed to investigate the conserved regions and epitopes among these viral proteins to validate the importance of the vaccines in Indonesian population.

## **2. MATERIALS AND METHODS**

### **2.1 Sample collection and rotavirus detection**

This study was a part of the Indonesian Rotavirus Surveillance study conducted in six teaching hospitals in Indonesia (Muhammad Hussein General Hospital, Palembang; Cipto Mangunkusumo General Hospital, Jakarta; Dr. Sardjito General Hospital, Yogyakarta; Hasan Sadikin General Hospital, Bandung; Sanglah General Hospital, Denpasar, Bali; and Mataram General Hospital, Mataram).<sup>13,14</sup> Fecal samples were

collected from children less than 5 years of age hospitalised due to acute diarrhea in the first 48 hours of admission based on the World Health Organisation (WHO).<sup>24</sup>

Fecal samples were transported to the Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta. All fecal samples were then aliquoted and stored at -80°C. The presence of group A rotavirus was detected by enzyme immunoassay (EIA), using IDEIA™ Rotavirus (DakoCytomation) kit according to the instructions provided by the manufacturer.

For the 2006 surveillance year, there were 2,293 stool samples collected and 1,245 of them (54.3%) were positive for rotavirus. For the 2009 surveillance year, 959 stool samples were collected and 456 of them (47.5%) were positive for rotavirus.<sup>13,14</sup>

## **2.2 Rotavirus genotyping**

Rotavirus RNA was extracted from rotavirus positive stool samples using the QIAamp RNA stool Minikit (Qiagen) according to the manufacturer instructions. Rotavirus RNA was subsequently analysed to determine both the VP7 (G-type) and the VP4 (P-type) genotypes using reverse transcriptase polymerase chain reaction (RT-PCR), as previously described.<sup>25-29</sup> The PCR products were first assigned to G- and P-genotype based on size comparison after electrophoresis on 2% agarose gel and ethidium bromide staining. The results of rotavirus detection and genotyping have been previously published.<sup>13,14</sup>

## **2.3 Sequence analysis, multiple alignments, and phylogenetic analysis**

Viral RNA was denatured by boiling at 97°C for 4 minutes and then transferred immediately to ice for 1 minute. The PCR reaction was conducted with a final volume of 30 µl containing 1.5 µl (20 µM) forward primer; 1.5 µl (20 µM) reverse primer; 1 µl

dNTP mix (10  $\mu$ M each); 5  $\mu$ l 5X buffer; 1  $\mu$ l one step RT-PCR enzyme mix (Qiagen); 16  $\mu$ l nuclease-free water; and 4  $\mu$ l RNA template. Reverse transcription was performed at 42°C for 30 minutes, 97°C for 15 minutes and followed by 35 cycles of 94°C for 30 seconds; 42°C for 30 seconds and 72°C for 60 seconds. The PCR product was electrophoresed on 1.2 % agarose gel and visualized with UV light.

The PCR products were purified using the Wizard SV Gel for PCR Clean-Up System (Promega, USA) according to the manufacturer's protocol, and then sequenced using an ABI PRISM BigDye Terminator Cycle Sequencing Reaction Kit (Applied Biosystems, Foster City, CA, USA) in an Applied Biosystems 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing of VP4, VP7, and NSP4 genes were conducted for selected samples with particular genotyping result, sampling location, and timing of sample collection. Electropherograms were visually analysed and contiguous DNA sequences were created utilizing the Sequencher® Software program version 5.0.1 (Gene Codes Corp Inc., An Arbor, MI, USA) for each gene segment.

We selected 20 strains for sequencing; 2, 4, 5, and 9 strains were G1P[6], G1P[8], G2P[4], and G9P[8], respectively. However, not all three genes of each strain were successfully sequenced. Therefore, in total we generated thirty seven sequences in this study, consisted of 16, 18, and 3 sequences of VP4, VP7, and NSP4 genes, respectively (Supplementary Table 1). All sequences obtained in this study were deposited in the GenBank under the accession numbers JQ837877 - JQ837884, JQ837886 - JQ837888 and MK814541-MK814567 (Supplementary Table 1).

The VP4, VP7, and NSP4 gene sequences from different regions of Indonesia were compared for homology with strains from other countries obtained from GenBank. For phylogenetic analysis, all complete and partial sequences from different Indonesian

regions were included to be compared with strains derived from other countries. The multiple sequence alignments were performed by using Muscle algorithm in MEGA version X.<sup>30</sup> The optimal evolutionary model was selected based upon the Akaike information criterion (corrected) (AICc) ranking implemented in jModelTest.<sup>31</sup> Maximum likelihood phylogenetic trees were generated using the selected models of nucleotide substitution: Hasegawa, Kishino and Yano (HKY+G<sub>G4</sub>) (G1 VP7), HKY+I (G2 VP7), General Time Reversible (GTR+G<sub>G4</sub>) (G9 VP7, P[4] VP4), GTR+I (P[8] VP4) and Tamura-Nei (TrN+G<sub>G4</sub>) (NSP4). The robustness of branches was assessed by bootstrap analysis using 1,000 pseudoreplicate runs and nodes with values >70% were considered to be strongly supported in the phylogenetic analysis. Nucleotide and amino acid distance matrixes were calculated using the p-distance algorithm in MEGA version 10.<sup>30</sup>

## **2.4 Immunoinformatic analysis**

For further comparative analysis of the sequenced genes, vaccine sequences of Rotarix and RotaTeq for the viral proteins VP4 (Rotarix: JN849113; RotaTeq: GU565044, GU565055, GU565066, GU565077, GU565088), VP7 (Rotarix: JN849114; RotaTeq: GU565057, GU565068, GU565079, GU565090, GU565046) and NSP4 (Rotarix; KX954623).<sup>32</sup> The amino acid sequence similarity of the vaccine strains was compared with Indonesian strains generated in this study. The antigenic regions of the Indonesian VP4, VP7 and NSP4 rotavirus proteins were also compared with the vaccine strains.<sup>33-35</sup>

### ***Analysis of conserved regions***

Conserved regions are known to be the best and the most effective drug and immune targets. Conservation analysis was performed among the rotavirus vaccine strains and

Indonesian rotavirus VP4, VP7 and NSP4 proteins (complete sequences) to determine the conserved regions comprised of  $\geq 9$  amino acid residues.<sup>36</sup>

### ***Epitope mapping***

For the comparative analysis of the putative antigenic regions of vaccine and Indonesian strains, epitope prediction was performed. The IEDB (Immune Epitope Database and Analysis Resource)<sup>37</sup> and ABCpred servers<sup>38</sup> were used for the prediction of B cell epitopes. ProPred-I<sup>39</sup> and ProPred<sup>40</sup> were utilized for the prediction of CD4+ (MHC II) and CD8+ (MHC I) T cell epitopes. ProPred-I and ProPred predict 47 types of MHC I and 57 types of MHC II allele specific binding peptides in a provided protein, respectively. Epitopes that bound to the maximum number of alleles were selected for further analysis. All predicted selected epitopes were further cross checked for homology with human proteomes by BLASTp<sup>41</sup> to prevent any possibility of autoimmune reactions.

## **2.5 Ethical approval**

This study has received the ethical clearance from the Ethical Committee of the Faculty of Medicine, Public Health and Nursing, Universitas Gajah Mada, Yogyakarta, Indonesia (KF/FK/519/EC).

## **3. RESULTS**

### **3.1 Phylogenetic analysis of VP4 gene of P[4] and P[8] Indonesian strains**

The VP4 gene of the three Indonesian G1P[8] strains were sequenced along with the VP4 gene of nine Indonesian G9P[8] strains. Phylogenetic analysis revealed the majority of strains formed multiple clusters within the same lineage. The G1P[8] strains and a single G9P[8] strain (RVA/Human-wt/IDN/MT-7346/2006/G9P8) clustered within a

sub-lineage interspersed with Indonesian equine-like G3P[8] strains circulating between 2014–2015 and contemporary global strains. Six G9P[8] strains formed multiple clusters within a sub-lineage comprised of global contemporary strains. A single strain, RVA/Human-wt/IDN/BL-5114/2006/G9P[8], clustered within a separate lineage closely related to the strain RVA/Human-wt/FIN/Fin-302/1988/G1P[8] (**Figure 1A**). The Indonesian strains from this study shared 87.2–100% nucleotide (nt) and 87.8–100% amino acid (aa) similarity.

The VP4 gene of four Indonesian G2P[4] strains were sequenced. The two strains from 2009 (RVA/Human-wt/IDN/YK-RVY47/2009/G2P[4] and RVA/Human-wt/IDN/MT-RVM113/2009/G2P[4]) clustered together, closely related to contemporary G2P[4] strains circulating in Australia and Taiwan. The two strains from 2006 (RVA/Human-wt/IDN/BL-5210/2006/G2P[4] and RVA/Human-wt/IDN/BL-5422/2006/GP2P[4]) clustered distinctly to the 2009 strains and were closely related to contemporary strains from Taiwan, Brazil, Belgium, Russia, and India (**Figure 1B**). The four Indonesian strains shared 97.3–99.6% nt and 97.9–100% aa similarity.

### **3.2 Phylogenetic analysis of VP7 gene of G1, G2, and G9 genotypes of Indonesian strains**

The VP7 genes of four Indonesian G1P[8] and two Indonesian G1P[6] strains sequenced in this study clustered within the same lineage and were closely related to contemporary strains from Japan, Thailand, and Vietnam (**Figure 2A**). The six Indonesian strains shared 97.7–100% nt and 98.4–100% aa similarity.

The VP7 gene of four Indonesian G2 strains sequenced in this study formed two sub-clusters. RVA/Human-wt/IDN/BL-5210/2006/G2P[4] and RVA/Human-wt/IDN/YK-

RVY47/2009/G2P[4] clustered with contemporary strains from Thailand. RVA/Human-wt/IDN/MT-RVM113/2009/G2P[4] and RVA/Human-wt/IDN/BD-4034/2006/G2P[4] clustered with contemporary strains from Australia, the Philippines, Taiwan and China (**Figure 2B**). The four Indonesian strains shared 92.8–99.9% nt and 95.2–99.6% aa similarity.

The VP7 gene of eight Indonesian G9P[8] strains sequenced in this study clustered within the same lineage with contemporary strains from Asia and Europe (**Figure 2C**). The eight Indonesian strains shared 98.0–100% nt and 98.0–100% aa similarity.

### 3.3 Phylogenetic analysis of NSP4 genes of Indonesian strains

Three NSP4 genes were sequenced from one G1P[8] strain and two G2P[4] strains (**Figure 3**). The G1P[8] strain (RVA/Human-wt/IDN/YK-RVY36/2009/G1P[8]) clustered within the E1 lineage closely related to strains from the USA and India. The two G2P[4] strains (RVA/Human-wt/IDN/MT-RVM113/2009/G2P[4] and JQ837886/RVA/Human-wt/IDN/BL-5210/2006/G2P[4]) clustered within the E2 lineage closely related to strains from India, Korea and Italy (**Figure 3**). The two E2 NSP4 genes shared 98.3% nt and 98.7% aa similarity.

### 3.4 Comparison of the VP4, VP7 and NSP4 antigenic regions of Indonesian and vaccine strains

The complete sequences of VP4, VP7 and NSP4 genes were obtained to compare the antigenic regions within these viral proteins with vaccine strains.<sup>33,34</sup> In VP4, there are four antigenic regions (8-1, 8-2, 8-3, 8-4)<sup>34</sup> (**Figure 4A**). When compared with RotaTaq and Rotarix vaccine strains (P8) four mutations were observed in the 8-1 region and five mutations in the 8-3 region. No mutations were observed in 8-2 and 8-4 regions.

Indonesian P4 strains exhibited two mutations in region 8-1, three mutations in region 8-3 but no mutations were observed in regions in 8-2 and 8-4 when the Indonesian strains were compared with the RotaTeq P8 strain (**Figure 4B**). Although there was a strain difference between Indonesian and vaccine (RotaTeq) strains, but interestingly only few variations were observed in 8-1 and 8-3 regions; while 8-2 and 8-4 regions were conserved. This region (VP4) also possesses fewer variations in comparison to the variations observed in P8 strains of Indonesia and vaccine strains.

In VP7 there are three antigenic regions (7-1a, 7-1b, 7-2)<sup>33</sup> (**Figure 4A**). When compared with vaccine strain (G1) seven mutations were observed in region 7-1a, five in region 7-1b and six in region 7-2. In Indonesian G2 strains only three variations were observed in 7-1b region when it was compared with the vaccine strain (G2) (**Figure 4B**).

In NSP4 three antigenic sections are present (ASI, ASII, ASIII) (**Figure 4A**).<sup>35</sup> When Indonesian strains were compared with the vaccine strain, seven variations were observed in ASI, ten in ASII, and two in ASIII (**Figure 4B**).

### 3.5 Conserved regions among VP4, VP7 and NSP4 genes

We have analysed all complete sequences of Indonesian VP4, VP7 and NSP4 and compared with RotaTeq and Rotarix vaccine strains to highlight all possible conserved regions in these viral proteins. In VP4, 27 conserved regions were found among the vaccine (Rotarix, RotaTeq) and P[8] Indonesian strains, while 14 conserved regions were observed in the vaccine and Indonesian strains (G2) (**Supplementary Table 2A**). In case of VP7, nine conserved regions were observed in the vaccine and P8 Indonesian strains, while ten conserved regions were found conserved between vaccine and Indonesian strain (G2) (**Supplementary Table 2B**). Indonesian NSP4 (G1) possessed only three conserved regions in comparison with vaccine strain, but

This article is protected by copyright. All rights reserved.

Indonesian NSP4 (G2) were shown to possess six conserved regions (**Supplementary Table 2C**). It was observed that despite of inter-genotypic differences, various conserved regions were common among the vaccine and Indonesian strains.

### 3.6 Epitope mapping of VP4, VP7, and NSP4 proteins

Epitope mapping of both vaccine and Indonesian strains was performed to determine B and T cell epitopes that are conserved among them and are present among the conserved regions. The conserved epitopes were further investigated for their antigenicity by using the online tool Vaxijen.<sup>42</sup> The epitopes were selected for their higher antigenicity i.e  $\geq 0.4$ . It was observed that most of the B cell epitopes lying within the conserved regions have more antigenicity conferring the importance of conserved regions as potential antigenic regions (**Figure 6**). Some of the conserved T cell epitopes (MHC I and II) also showed high antigenicity (**Figure 5**), thus showing the potential of these regions. The details of epitope sequence were provided in Supplementary Table 3.

## 4. DISCUSSION

Rotavirus is known to be a major cause of severe diarrhea globally. The World Health Organisation (WHO) has recommended rotavirus vaccines to be included in worldwide through national immunisation programs.<sup>43</sup> It has been reported that rotavirus infection accounts for a substantial disease burden in children less than 5 years of age in Indonesia.<sup>14</sup> However, rotavirus vaccines have not been introduced in the national immunisation program of Indonesia. The decision to introduce rotavirus vaccine in Indonesia is delayed due to the lack of awareness of the importance of rotavirus vaccines as well as financial constraints of the government.<sup>44</sup> In this report, we performed an investigational study of circulating rotavirus strains and immunoinformatic

comparison with the vaccine strains to raise awareness of rotavirus vaccine introduction in Indonesia.

G1P[8], G2P[4], and G9P[8] strains were dominant during this study. We observed that rotavirus genotypes circulating in Indonesia were similar in comparison to global strains.<sup>17</sup> It was recognized that G1P[8], G2P[8], and G9P[8] were the dominant genotypes of circulating strains in Indonesia during 2006-2009, which were also reported in previous global studies.<sup>17</sup> Within each genotype, strains detected in 2006-2009 clustered in a single common lineage and were closely related to contemporary strains detected globally. However, it was suggested that the predominant rotavirus genotypes in Indonesia vary by year and location, underlying the importance of a continuous investigation on circulating genotypes.<sup>13,45</sup>

One mechanism of generating genetic diversity is accumulation of point mutations (genetic drift) in each replication cycle of circulating rotavirus strains.<sup>34</sup> In the present study, we compared amino acid sequences of the VP4, VP7, and NSP4 antigenic domains of Indonesian and vaccine strains to determine the possible variations that could affect vaccine efficacy. We observed that certain variations were present among Indonesian and vaccine strains. These variations were distributed heterogeneously on the surface of the protein (**Figure 4**).

Genetic heterogeneity of various rotavirus genotypes could affect the immune responses against the vaccine.<sup>46</sup> Hence, more specific and conserved epitopes need to be targeted in order to design a highly protective vaccine against circulating rotavirus genotypes. Furthermore, knowledge of currently circulating strains is required before the initiation of a vaccination program in Indonesia. The most appropriate vaccine design requires information about distribution of pathogen genotypes as well as HLA alleles.<sup>47</sup>

Through this method, more specific epitopes could be selected which can be valuable to design epitope-based vaccines to efficiently elicit potent B and T cell response.<sup>48</sup> Based on these observations, we further determined the conserved regions among the vaccine and Indonesian strains. Further B and T cell epitope prediction was performed for both vaccine and Indonesian strains. Only those epitopes that laid in the conserved regions and have high antigenicity were then selected. Our analysis showed that there are a wide variety of conserved epitopes that are common between vaccine and Indonesian strains, conferring the importance of the available rotavirus vaccine in Indonesian population.

Thus, our present study describes comparison between the rotavirus strains circulating in Indonesia with Rotarix and RotaTeq vaccine strains as well as with global strains. The genotyping and immunoinformatic analysis from the present data provides necessary information to identify novel and emerging rotavirus strains but also immunoinformatic strategy for vaccine development. However, they should be further validated in functional assays of cross protection experiments or field-based analysis to prove the biological relevance of our findings.

## **5. CONCLUSIONS**

The present study provides comprehensive data about the genetic diversity of circulating rotavirus strains in several regions in Indonesia. Continuous surveillance is essential to monitor circulating genotypes and emergence of novel variants as well as to predict the vaccine efficacy both before and after introduction of mass vaccination.

## **ACKNOWLEDGEMENT**

This research was supported by the grant from the Ministry of Health, Indonesia (RISBINIPTEKDOK 2009/2010). We thank Murdoch Children Research Institute Melbourne, Australia, for providing facilities to conduct genotyping and sequencing; Daniel Cowley and Nada Bogdanovic-Sakran for their technical support; and the parents of the children involved in this study for their willingness to participate. The sequencing data of this study was obtained from the dissertation of Hera Nirwati.

## **CONFLICTS OF INTEREST**

The authors declare that there are no conflicts of interest.

## **AUTHORS' CONTRIBUTIONS**

H. N., C. M. D., and A. I. contributed to the study concept and design, acquisition of data, analysis and interpretation of data, and drafting the manuscript; T. W. and A. T. A. contributed to acquisition and interpretation of data as well as critical revision of the manuscript; C. D. K., Y. S. and Q. P. contributed to the scientific discussion, provide facilities, and critical revision of the manuscript; M. S. H. contributed to study concept and design, study supervision, finalisation and critical revision of the manuscript.

## **ORCID**

*Mohamad S. Hakim* <http://orcid.org/0000-0001-8341-461X>

**REFERENCES**

1. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 2010; 375(9730): 1969-87.
2. Walker CL, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet* 2013; 381(9875): 1405-16.
3. Lanata CF, Fischer-Walker CL, Olascoaga AC, et al. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. *PLoS One* 2013; 8(9): e72788.
4. Tate JE, Burton AH, Boschi-Pinto C, Parashar UD, World Health Organization-Coordinated Global Rotavirus Surveillance. Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000-2013. *Clin Infect Dis* 2016; 62 Suppl 2: S96-S105.
5. Greenberg HB, Estes MK. Rotaviruses: from pathogenesis to vaccination. *Gastroenterology* 2009; 136(6): 1939-51.
6. Gentsch JR, Laird AR, Bielfelt B, et al. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *J Infect Dis* 2005; 192 Suppl 1: S146-59.
7. Iturriza-Gomara M, Isherwood B, Desselberger U, Gray J. Reassortment in vivo: driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. *J Virol* 2001; 75(8): 3696-705.
8. Tacharoenmuang R, Komoto S, Guntapong R, et al. Whole genomic analysis of an unusual human G6P[14] rotavirus strain isolated from a child with diarrhea in Thailand: evidence for bovine-to-human interspecies transmission and reassortment events. *PLoS One* 2015; 10(9): e0139381.
9. Tam KI, Roy S, Esona MD, et al. Full genomic characterization of a novel genotype combination, G4P[14], of a human rotavirus strain from Barbados. *Infect Genet Evol* 2014; 28: 524-9.
10. Ward ML, Mijatovic-Rustempasic S, Roy S, et al. Molecular characterization of the first G24P[14] rotavirus strain detected in humans. *Infect Genet Evol* 2016; 43: 338-42.
11. Dennis FE, Fujii Y, Haga K, et al. Identification of novel Ghanaian G8P[6] human-bovine reassortant rotavirus strain by next generation sequencing. *PLoS One* 2014; 9(6): e100699.

12. Kawai K, O'Brien MA, Gouveia MG, Mast TC, El Khoury AC. Burden of rotavirus gastroenteritis and distribution of rotavirus strains in Asia: a systematic review. *Vaccine* 2012; 30(7): 1244-54.
13. Nirwati H, Wibawa T, Aman AT, Wahab A, Soenarto Y. Detection of group A rotavirus strains circulating among children with acute diarrhea in Indonesia. *Springerplus* 2016; 5: 97.
14. Soenarto Y, Aman AT, Bakri A, et al. Burden of severe rotavirus diarrhea in Indonesia. *J Infect Dis* 2009; 200 Suppl 1: S188-94.
15. Wilopo SA, Soenarto Y, Bresee JS, et al. Rotavirus surveillance to determine disease burden and epidemiology in Java, Indonesia, August 2001 through April 2004. *Vaccine* 2009; 27 Suppl 5: F61-6.
16. Mulyani NS, Prasetyo D, Karyana IPG, et al. Diarrhea among hospitalized children under five: a call for inclusion of rotavirus vaccine to the national immunization program in Indonesia. *Vaccine* 2018; 36(51): 7826-31.
17. Banyai K, Laszlo B, Duque J, et al. Systematic review of regional and temporal trends in global rotavirus strain diversity in the pre rotavirus vaccine era: insights for understanding the impact of rotavirus vaccination programs. *Vaccine* 2012; 30 Suppl 1: A122-30.
18. Dennehy PH. Rotavirus vaccines: an overview. *Clin Microbiol Rev* 2008; 21(1): 198-208.
19. Desai R, Oliveira LH, Parashar UD, Lopman B, Tate JE, Patel MM. Reduction in morbidity and mortality from childhood diarrhoeal disease after species A rotavirus vaccine introduction in Latin America - a review. *Mem Inst Oswaldo Cruz* 2011; 106(8): 907-11.
20. Parashar U, Steele D, Neuzil K, et al. Progress with rotavirus vaccines: summary of the Tenth International Rotavirus Symposium. *Expert Rev Vaccines* 2013; 12(2): 113-7.
21. Armah GE, Sow SO, Breiman RF, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. *Lancet* 2010; 376(9741): 606-14.
22. Zaman K, Dang DA, Victor JC, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. *Lancet* 2010; 376(9741): 615-23.

23. Jiang V, Jiang B, Tate J, Parashar UD, Patel MM. Performance of rotavirus vaccines in developed and developing countries. *Hum Vaccin* 2010; 6(7): 532-42.
24. World Health Organization (2002) Generic protocol for (i) hospital-based surveillance to estimate the burden of rotavirus gastroenteritis in children and (ii) a community-based survey on utilization of health care services for gastroenteritis in children. Field test version. Geneva, Switzerland: Department of Vaccines and Biologicals, World Health Organization.
25. Das BK, Gentsch JR, Cicirello HG, et al. Characterization of rotavirus strains from newborns in New Delhi, India. *J Clin Microbiol* 1994; 32(7): 1820-2.
26. Gomara MI, Cubitt D, Desselberger U, Gray J. Amino acid substitution within the VP7 protein of G2 rotavirus strains associated with failure to serotype. *J Clin Microbiol* 2001; 39(10): 3796-8.
27. Gouvea V, Glass RI, Woods P, et al. Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. *J Clin Microbiol* 1990; 28(2): 276-82.
28. Gentsch JR, Glass RI, Woods P, et al. Identification of group A rotavirus gene 4 types by polymerase chain reaction. *J Clin Microbiol* 1992; 30(6): 1365-73.
29. Simmonds MK, Armah G, Asmah R, et al. New oligonucleotide primers for P-typing of rotavirus strains: Strategies for typing previously untypeable strains. *J Clin Virol* 2008; 42(4): 368-73.
30. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018; 35(6): 1547-9.
31. Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 2008; 25(7): 1253-6.
32. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2014; 42(Database issue): D32-7.
33. Aoki ST, Settembre EC, Trask SD, Greenberg HB, Harrison SC, Dormitzer PR. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science* 2009; 324(5933): 1444-7.
34. Zeller M, Patton JT, Heylen E, et al. Genetic analyses reveal differences in the VP7 and VP4 antigenic epitopes between human rotaviruses circulating in Belgium and rotaviruses in Rotarix and RotaTeq. *J Clin Microbiol* 2012; 50(3): 966-76.

35. Tatsumi M, Nagaoka Y, Tsugawa T, Yoto Y, Hori T, Tsutsumi H. Characterization of the NSP4 gene of group A human rotavirus G1P[8] strains circulating in Sapporo, Japan from 1987 to 2000. *J Med Virol* 2014; 86(2): 354-9.
36. Suri A, Walters JJ, Levisetti MG, Gross ML, Unanue ER. Identification of naturally processed peptides bound to the class I MHC molecule H-2Kd of normal and TAP-deficient cells. *Eur J Immunol* 2006; 36(3): 544-57.
37. Vita R, Overton JA, Greenbaum JA, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 2015; 43(Database issue): D405-12.
38. Saha S, Raghava GP. Prediction methods for B-cell epitopes. *Methods Mol Biol* 2007; 409: 387-94.
39. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 2003; 19(8): 1009-14.
40. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 2001; 17(12): 1236-7.
41. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008; 36(Web Server issue): W5-9.
42. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 2007; 8: 4.
43. Tate JE, Burton AH, Boschi-Pinto C, et al. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis* 2012; 12(2): 136-41.
44. Suwantika AA, Zakiyah N, Lestari K, Postma MJ. Accelerating the introduction of rotavirus immunization in Indonesia. *Expert Rev Vaccines* 2014; 13(4): 463-72.
45. Hakim MS, Nirwati H, Aman AT, Soenarto Y, Pan Q. Significance of continuous rotavirus and norovirus surveillance in Indonesia. *World J Pediatr* 2018; 14(1): 4-12.
46. Matthijssens J, Bilcke J, Ciarlet M, et al. Rotavirus disease and vaccination: impact on genotype diversity. *Future Microbiol* 2009; 4(10): 1303-16.
47. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev* 2009; 22(2): 370-85, Table of Contents.
48. Kumai T, Fan A, Harabuchi Y, Celis E. Cancer immunotherapy: moving forward with peptide T cell vaccines. *Curr Opin Immunol* 2017; 47: 57-63.

**FIGURES**

Figure 1. Maximum likelihood phylogenetic analysis of the VP4 gene of Indonesian P[8] (A); and P[4] (B) strains. Strains characterised in this study (marked by ■) were analysed with global isolates obtained from GenBank. Bootstrap values  $\geq 70\%$  are shown. Scale bar shows substitutions per nucleotide.

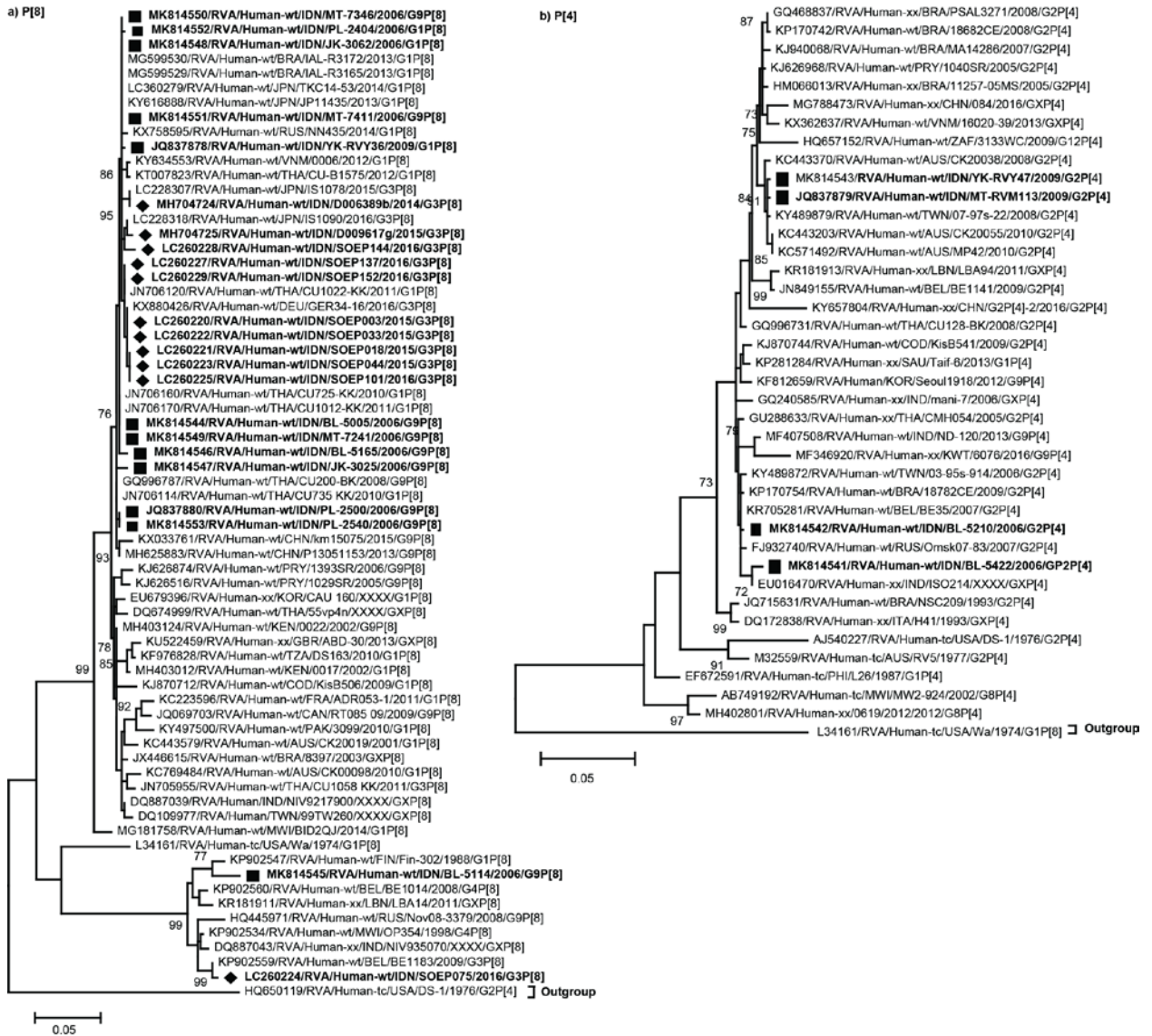


Figure 2. Maximum likelihood phylogenetic analysis of the VP7 gene of Indonesian G1 (A); G2 (B); and G9 (C) strains. Strains characterised in this study (marked by ■), and Indonesian strains from GenBank (marked by ◆) were analysed with global isolates obtained from GenBank. Bootstrap values  $\geq 70\%$  are shown. Scale bar shows substitutions per nucleotide.

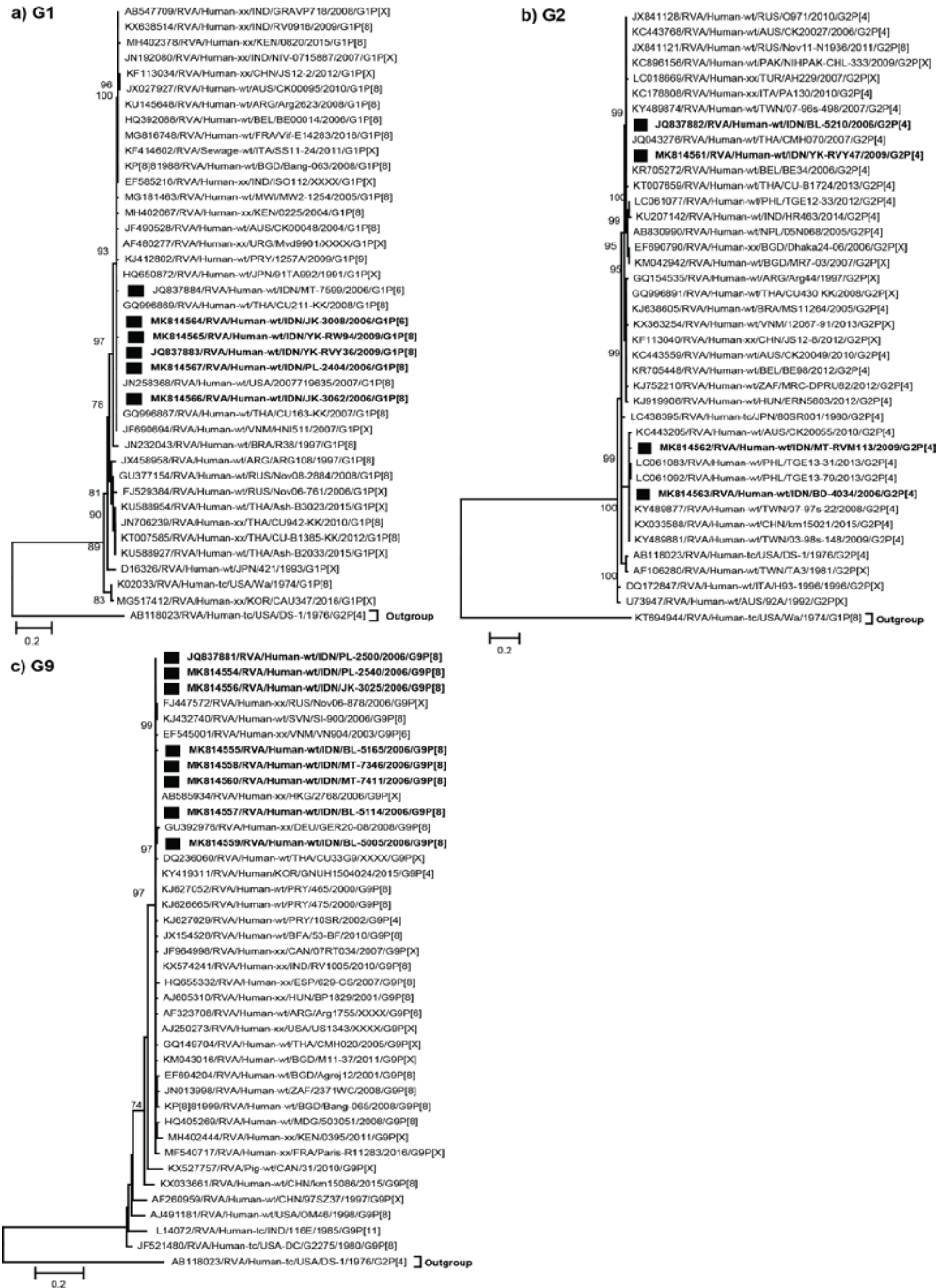


Figure 3. Maximum likelihood phylogenetic analysis of the NSP4 gene of Indonesian isolates. Strains characterised in this study (marked by ■), and Indonesian strains from GenBank (marked by ◆) were analysed with global isolates obtained from GenBank. Bootstrap values  $\geq 70\%$  are shown. Scale bar shows substitutions per nucleotide.

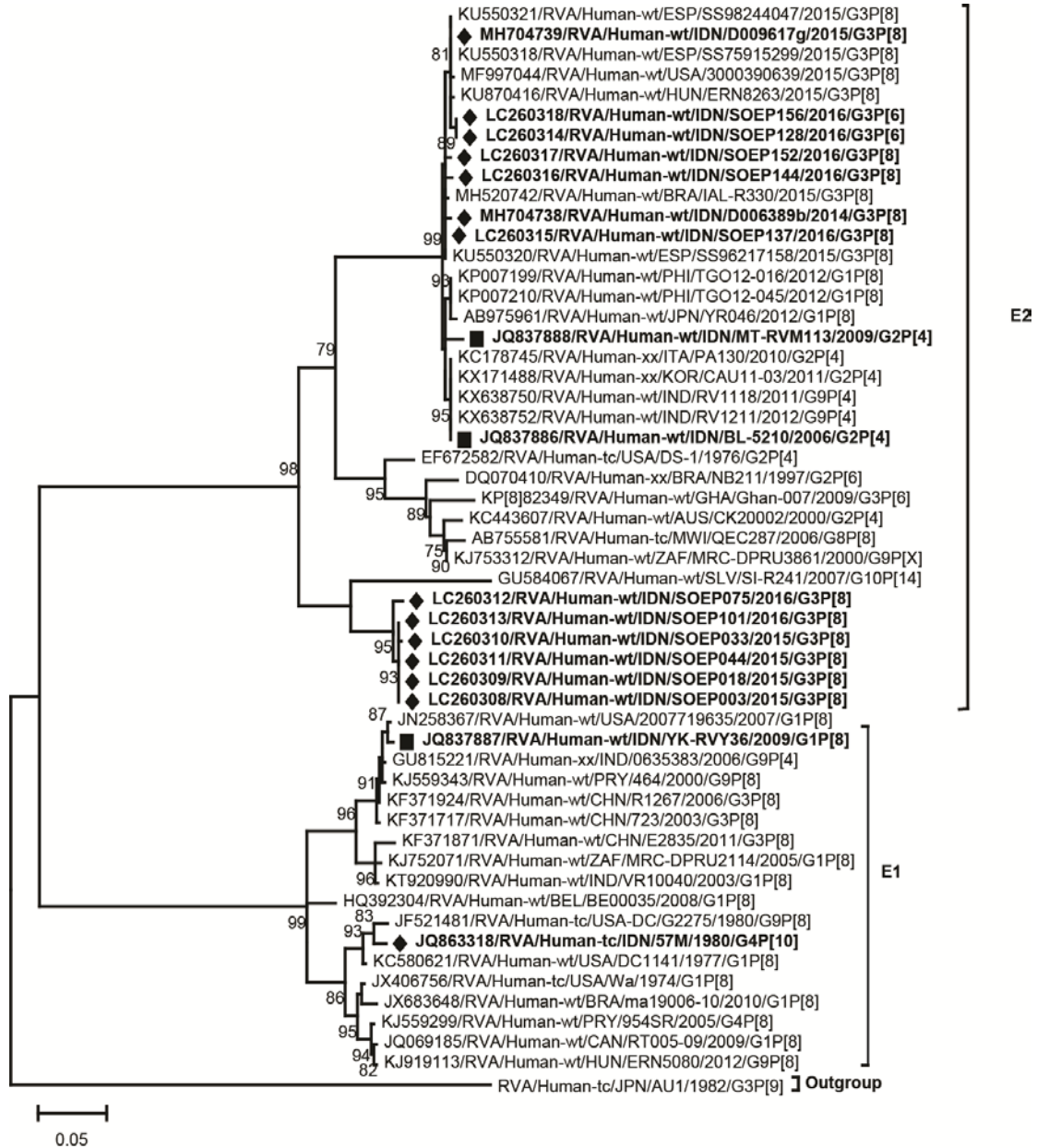


Figure 4. (A) Alignment of antigenic residues in VP4, VP7 and NSP4 between the strains contained in Rotarix, RotaTeq and Indonesian circulating strain. Antigenic residues are divided in four antigenic epitopes (8-1, 8-2, 8-3, 8-4) in VP4 protein; three antigenic epitopes (7-1a, 7-1b, and 7-2) in VP7; and three antigenic epitopes (ASI, ASII and ASIII) in NSP4 protein. Amino acid residues that differ between vaccine and Indonesian strains are indicated in colored boxes a black triangle. (B) Surface representation of VP4, VP7 and NSP4 proteins. Antigenic epitopes are colored in red (8-1), yellow (8-2), green (8-3), and purple (8-4) in VP4; red (7-1a), green (7-1b), and purple (7-2) in VP7; and red (ASI), blue (ASII) and green (ASIII) in NSP4. Only the surface showing the antigenic epitopes are presented by rotating at 180°.

| Strains         | VP4  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |    |    |  |  |  |
|-----------------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|--|--|--|
|                 | 08_1 |     |     |     |     |     |     |     | 8_2 |     |     |     | 8_3 |     |     |     |     |     |     |     | 8_4 |     |    |    |    |  |  |  |
|                 | 100  | 146 | 148 | 150 | 188 | 190 | 192 | 193 | 194 | 195 | 196 | 180 | 183 | 113 | 114 | 115 | 116 | 125 | 131 | 132 | 133 | 135 | 87 | 88 | 89 |  |  |  |
| ROTARIX         | D    | S   | Q   | E   | S   | T   | N   | L   | N   | N   | I   | T   | A   | N   | P   | V   | D   | S   | S   | N   | D   | N   | N  | T  | N  |  |  |  |
| ROTATEQ         | *    | *   | *   | *   | *   | *   | *   | *   | *   | D   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *  | *  | *  |  |  |  |
| YK-RVY36_G1P[8] | *    | A   | *   | D   | *   | *   | *   | *   | *   | G   | *   | *   | *   | D   | S   | *   | *   | N   | R   | *   | *   | D   | *  | *  |    |  |  |  |
| PL-2500_G9P[8]  | *    | A   | *   | *   | *   | *   | *   | *   | *   | G   | *   | *   | *   | D   | *   | *   | *   | N   | R   | *   | *   | *   | *  | *  |    |  |  |  |
| ROTATEQ         | G    | S   | D   | D   | Y   | V   | N   | Y   | A   | S   | V   | T   | A   | T   | S   | E   | T   | S   | S   | N   | A   | D   | T  | G  | P  |  |  |  |
| RVM113_G2P[4]   | *    | A   | *   | *   | *   | *   | *   | *   | *   | G   | *   | *   | *   | D   | *   | *   | *   | N   | R   | *   | *   | *   | *  | *  |    |  |  |  |
| BL5210_G2P[4]   | *    | A   | *   | *   | *   | *   | *   | *   | *   | G   | *   | *   | *   | D   | *   | *   | *   | N   | R   | *   | *   | *   | *  | *  |    |  |  |  |

| Strains           | VP7  |    |    |    |    |    |    |     |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-------------------|------|----|----|----|----|----|----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                   | 7_1a |    |    |    |    |    |    |     | 7_1b |     |     |     |     |     |     |     | 7_2 |     |     |     |     |     |     |     |     |     |     |     |     |
|                   | 87   | 91 | 94 | 96 | 97 | 98 | 99 | 100 | 104  | 123 | 125 | 129 | 130 | 291 | 201 | 211 | 212 | 213 | 238 | 242 | 143 | 145 | 146 | 147 | 148 | 190 | 217 | 221 | 264 |
| ROTARIX           | T    | T  | N  | G  | E  | W  | K  | D   | Q    | S   | V   | V   | D   | K   | Q   | N   | V   | D   | N   | T   | K   | D   | Q   | N   | L   | S   | M   | N   | G   |
| ROTATEQ           | *    | *  | *  | D  | *  | *  | *  | *   | *    | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | S   | *   | *   | *   | *   | *   | *   |
| YK-RVY36 (G1P[8]) | *    | *  | S  | *  | *  | *  | *  | *   | N    | A   | I   | *   | *   | *   | *   | D   | *   | A   | D   | N   | *   | *   | S   | T   | *   | *   | E   | S   | *   |
| PL-2500 (G9P[8])  | *    | *  | G  | T  | *  | *  | *  | *   | D    | A   | I   | *   | *   | *   | *   | *   | *   | A   | D   | N   | *   | *   | S   | T   | *   | *   | E   | S   | *   |
| ROTATEQ           | A    | N  | S  | D  | E  | W  | E  | N   | Q    | D   | T   | M   | N   | K   | Q   | D   | V   | S   | N   | S   | R   | D   | N   | T   | L   | D   | I   | S   | G   |
| BL-5210 (G2P[4])  | *    | *  | *  | *  | *  | *  | *  | *   | *    | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | N   | *   | *   | *   | *   | *   | *   | *   | *   | *   |

| Strains          | NSP4  |     |     |     |     |     |     |     |     |     |     |     |     |     |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
|------------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
|                  | ASIII |     |     |     |     |     |     |     |     |     |     |     |     |     | ASII |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
|                  | 112   | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126  | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |   |
| ROTARIX          | M     | I   | D   | K   | L   | T   | T   | R   | E   | I   | E   | Q   | V   | E   | L    | L   | K   | R   | I   | H   | D   | N   | T   | R   | P   | V   | D   | V   | I   | D   | M   | S   | K   | E   | F   | N   | Q   |   |
| YK-RVY36_G1P[8]  | *     | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *    | *   | *   | *   | *   | *   | *   | *   | *   | *   | G   | Q   | S   | I   | G   | E   | *   | *   | T   | *   | *   | *   | *   | * |
| BL-5210_G2P[4]   | *     | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *    | *   | *   | *   | *   | *   | *   | *   | Y   | K   | V   | Q   | S   | I   | G   | E   | *   | *   | T   | *   | *   | *   | *   |   |
| MT-RVM113_G2P[4] | *     | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *   | *    | *   | *   | *   | *   | *   | *   | *   | Y   | K   | V   | Q   | S   | I   | G   | E   | *   | *   | T   | *   | *   | *   | *   |   |

| Strains          | ASI |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                  | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 |
| ROTARIX          | N   | I   | K   | T   | L   | D   | E   | W   | E   | S   | G   | K   | N   | P   | Y   | E   | P   | S   |
| YK-RVY36_G1P[8]  | *   | *   | *   | *   | *   | *   | *   | *   | *   | N   | *   | *   | *   | *   | *   | *   | *   | I   |
| BL-5210_G2P[4]   | V   | R   | *   | *   | E   | *   | *   | *   | *   | *   | R   | *   | *   | *   | *   | *   | *   | K   |
| MT-RVM113_G2P[4] | V   | R   | *   | *   | E   | *   | *   | *   | *   | *   | R   | *   | *   | *   | *   | *   | *   | K   |

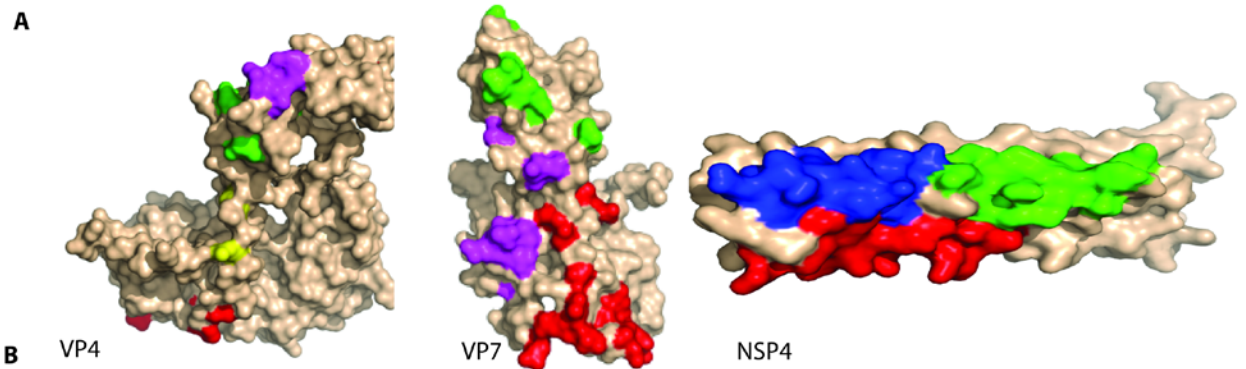


Figure 5. Antigenicity of the conserved T cell (MHC I and MHC II) epitopes among the vaccines and Indonesian strains.

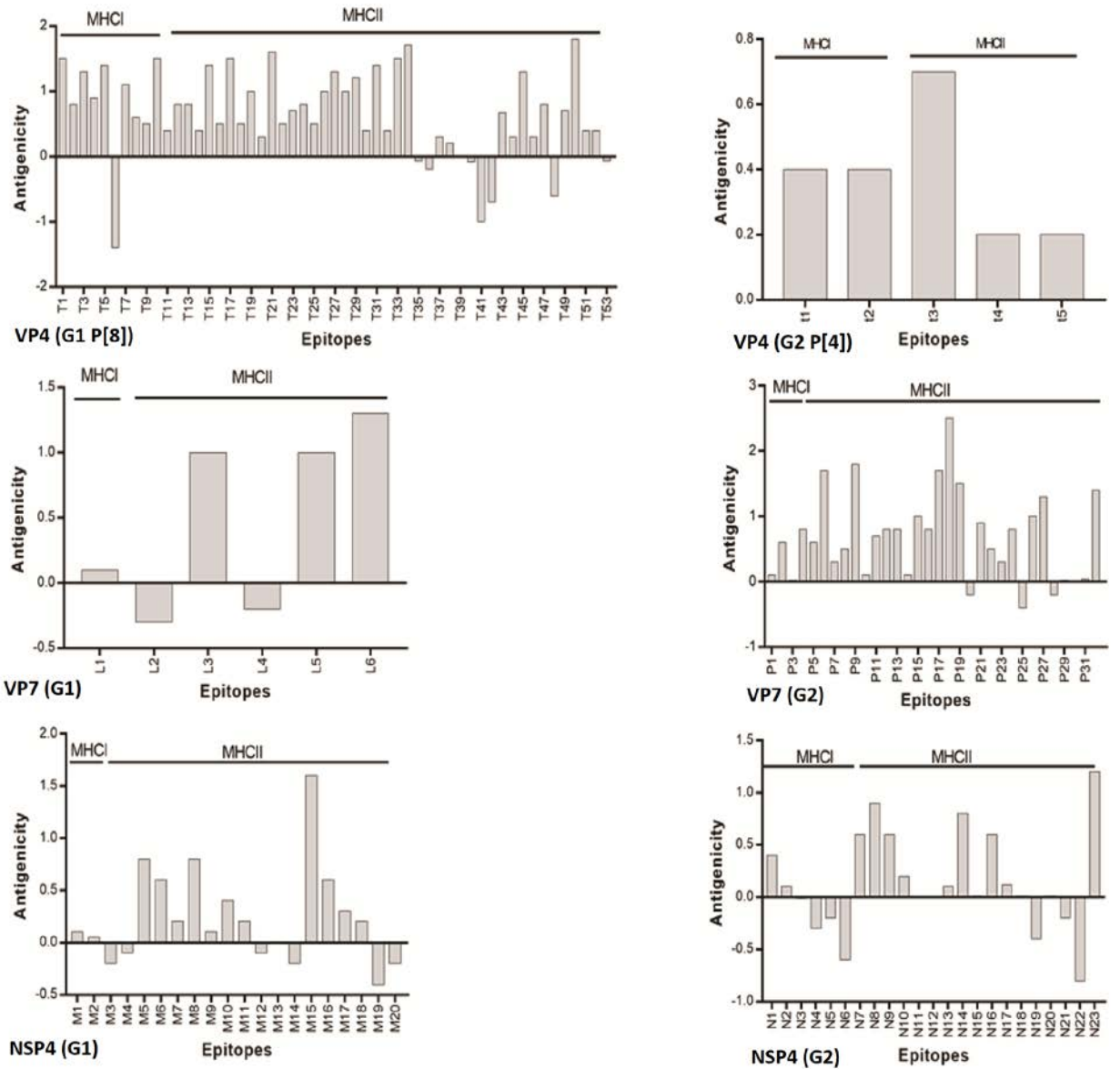


Figure 6. Antigenicity of the conserved B cell epitopes among the vaccines and Indonesian strains.

