Future zoonosis forecasting of Betacoronavirus group using phylogenetics and haplotype network analysis

Rahmat Grahadi1, Rizka Vamelia Sulistyia Ningrum1, Najma Zahira1, Nia Kurniawan1,2, Fatchiyah Fatchiyah1,3

1Department of Biology, Faculty of Mathematics and Natural Science, Universitas Brawijaya, Malang, Indonesia
2NK Research, Faculty Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia
3Research Center of Smart Molecule of Natural Genetics Resources, Universitas Brawijaya, Malang, Indonesia

Abstract

Genetic diversity and species relationship contribute to the inter-species transmission of coronavirus-derived diseases. This research aimed to investigate the probability of newly emergence Betacoronavirus zoonosis according to genetic relationship and haplotype network. A total of 24 Betacoronavirus sequences from different continents were used to construct the phylogenetic tree and the haplotype network. A phylogenetic tree was constructed using the Maximum Likelihood method with RaxML in CIPRES Science Gateway portal and GTRGAMMA+I analysis with 1000 bootstrap. Haplotype Network was done using Network 10.0.0.0 version with median-joining analysis. The result of the phylogenetic and haplotype network formed 4 groups of Betacoronavirus. Group A consisted of Human Coronavirus (HCoV) type OC43 and HKU1, Bovine Coronavirus, and Rodent Coronavirus. Group B consisted of SARS-CoV and SARS-CoV-2. Group C consisted of MERS-CoV (2012 pandemic) and MERS-CoV from camels. Reliable with phylogenetic results, haplotype network also grouped Betacoronaviruses into three groups in accordance with their subgenera. Bovine and Rodent coronavirus are constant to group with previously human coronavirus, i.e. HCoV-OC43 and HCoV-HKU1, respectively. According to the high genetic similarity that the Bovine and Rodent coronavirus may infect to human and provide a new emergence. This study is basic for further research related to inter-species transmission from animal to human.

Keywords: Betacoronavirus, haplotype network, phylogenetic, zoonosis

Received: 8 June 2020 Revised: 13 June 2020 Accepted: 15 June 2020

Introduction

A new Coronavirus (CoV) became an outbreak since 2019 and infected more than 2 million people in the world. The virus originally came from Wuhan, China then becomes a mass outbreak to other countries. After identified, the virus was named as SARS-CoV-2 with Coronavirus Disease-19 (COVID-19) as it caused illness (Du et al., 2020). Later on, it is known that SARS-CoV-2 belongs to Betacoronavirus, which includes MERS (Middle East Respiratory Syndrome) and SARS (Severe Acute Respiratory Syndrome) (Wassenaar & Zou, 2020).

Coronavirus is a member of the order Nidovirales, family Coronaviridae, and subfamily Coronavirinae. Coronavirinae then divided into 4 genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus. CoV has an envelope and also contains a positive sense RNA around 32 kb length. Structurally, CoV contains envelope protein, membrane protein, nucleocapsid protein, and spike protein (Chen et al., 2020; Cui et al., 2019). The genome of common CoV contains no less than six Open Reading Frames (ORF), with ORF1 encodes non-structural transcription-replication complex, while later ORFs encode for viral structural protein like Membrane (M), Envelope (E), and Spike (S) proteins (Chen et al., 2020).

Before the COVID-19 outbreak, the emergence of zoonotic diseases transmitted from the animal reservoirs has become a serious threat to public health. For instance, in Betacoronavirus cases only, SARS-CoV and MERS-CoV have been declared as major health problems after infecting many people across the world with high mortality (de Wit et al., 2016). On the other hand, genetic diversity among transmitted viruses from animals was believed to contribute to the zoonosis process (Ji et al., 2020; Lu et al., 2018; Woo et al., 2009). Hence, observing previously known animal-derived coronaviruses will provide an insight for predicting future transmission from the original host into human. This study will give some awareness of the possibility of forthcoming coronavirus zoonosis based on genetic similarity and viral species relationship.

Methods

Data Mining

Twenty-one whole genomes sequences of Betacoronavirus from several countries in the world were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) using a random sampling method to ensure the variability (Yang et al., 2014). Besides, we add all of the SARS-CoV-2
Indonesian genome (3 genomes, per 6 May 2020) from variances with other countries. This set of data consist of all species of Betacoronavirus and Canine coronavirus from Alphacoronavirus for outgroup (Tab. 1).

**Data Analysis**

The sequences were aligned using ClustalW method at CIPRES Science Gateway portal. Polymorphism and gene flow analysis among SARS-CoV-2 sequences was analyzed by using DNASP v.6.12.03 (Rozas et al., 2017) to compare its genetic diversity and gene flow between China population and other countries. Gene flow strength (Nm) uses the Nei (1973) criterion formula (Nm = [1/Gst−1]), with Nm<1 considered low gene flow and Nm>1 considered high gene flow (Hwang & Cho, GISAID (https://www.gisaid.org/) to compare genome 2018). The phylogenetic tree was constructed using Maximum Likelihood method with RaxML in CIPRES Science Gateway Portal (Stamatakis, 2014). Phylogenetic analysis for Maximum Likelihood uses GTR+GAMMA+I algorithms with 1000 bootstrap, the higher bootstrap value more accurate the phylogenetic. The result was visualized using FigTree v1.4.4 and saved in jpg format (Rambaut, 2018). Haplotype Network was conducted using Network 10.0.0.0 version (Fluxus Technology, 2020) with median-joining analysis (Bandelt et al., 1999). NCBI BLAST carried out to analyze potential zoonotic, high genetic similarity more likely transfer to human (NCBI, 2020).

**Table 1.** The identity of used sequences in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>Species</th>
<th>Location</th>
<th>Accession Number</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SARS-CoV</td>
<td>China</td>
<td>KY369905.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>2.</td>
<td>SARS-CoV</td>
<td>Korea</td>
<td>LC522975.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>3.</td>
<td>SARS-CoV</td>
<td>Italy</td>
<td>YP008712.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>4.</td>
<td>SARS-CoV</td>
<td>Canada</td>
<td>NC_004718</td>
<td>He et al., 2004</td>
</tr>
<tr>
<td>5.</td>
<td>SARS-CoV</td>
<td>Germany</td>
<td>YP008712.1</td>
<td>He et al., 2004</td>
</tr>
<tr>
<td>6.</td>
<td>SARS-CoV-2</td>
<td>Japan</td>
<td>KT125476.2</td>
<td>Piplat et al, 2017</td>
</tr>
<tr>
<td>7.</td>
<td>SARS-CoV-2</td>
<td>China: Shenzhen</td>
<td>MN938384.1</td>
<td>Chan et al., 2020</td>
</tr>
<tr>
<td>8.</td>
<td>SARS-CoV-2</td>
<td>Korea</td>
<td>MT039890.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>9.</td>
<td>SARS-CoV-2</td>
<td>USA: California</td>
<td>MN994468.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>10.</td>
<td>SARS-CoV-2</td>
<td>Sweden</td>
<td>MT093571.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>11.</td>
<td>SARS-CoV-2</td>
<td>USA: Wuhan</td>
<td>MN908947.3</td>
<td>Wu et al., 2020</td>
</tr>
<tr>
<td>12.</td>
<td>SARS-CoV-2</td>
<td>USA: Illinois</td>
<td>MN988713.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>13.</td>
<td>MERS-CoV</td>
<td>Kenya</td>
<td>MH734115.1</td>
<td>Omeh et al, 2018</td>
</tr>
<tr>
<td>14.</td>
<td>MERS-CoV</td>
<td>France</td>
<td>KJ361503.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>15.</td>
<td>MERS-CoV</td>
<td>Thailand</td>
<td>KT225476.2</td>
<td>Piplat et al, 2017</td>
</tr>
<tr>
<td>16.</td>
<td>MERS-CoV</td>
<td>China</td>
<td>KT006149.2</td>
<td>Lu et al, 2015</td>
</tr>
<tr>
<td>17.</td>
<td>HCoV-OC43</td>
<td>USA</td>
<td>KJ361503.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>18.</td>
<td>Bovine Coronavirus</td>
<td>Japan</td>
<td>LC49138.1</td>
<td>Suzuki et al, 2020</td>
</tr>
<tr>
<td>19.</td>
<td>HCoV-HKU1</td>
<td>Thailand</td>
<td>MH940245.1</td>
<td>Unpublished</td>
</tr>
<tr>
<td>20.</td>
<td>Rodent Coronavirus</td>
<td>China</td>
<td>KY370043.1</td>
<td>Wu et al, 2018</td>
</tr>
<tr>
<td>21.</td>
<td>Canine Coronavirus</td>
<td>Italy</td>
<td>KP981644.1</td>
<td>Decarbo et al, 2015</td>
</tr>
<tr>
<td>22.</td>
<td>SARS-CoV-2</td>
<td>Indonesia: Jakarta 1</td>
<td>EPI_ISL_435281*</td>
<td>Unpublished</td>
</tr>
<tr>
<td>23.</td>
<td>SARS-CoV-2</td>
<td>Indonesia: Jakarta 2</td>
<td>EPI_ISL_435282*</td>
<td>Unpublished</td>
</tr>
<tr>
<td>24.</td>
<td>SARS-CoV-2</td>
<td>Indonesia: Jakarta 3</td>
<td>EPI_ISL_435283*</td>
<td>Unpublished</td>
</tr>
</tbody>
</table>

**Results**

**Alignment and Phylogenetic**

According to alignment result, HCoV-OC43, Bovine CoV, HCoV-HKU1, Rodent CoV, SARS-CoV, SARS-CoV-2, MERS-CoV, and also Canine CoV showed their distinguishing feature. With a focus on SARS-CoV-2, polymorphic analysis shows there were 29-mutation sites discovered, locating on the ORF1a (Base number 1714, 2767, 2822, 3091, 6381, 6957, 9327, 9828, 11661, 12732, and 13446), ORF1b (13849, 13850, 15275, 16236, 17643, and 18019), S protein (21705, 21706, 22692, 23277, 25255, 27334), ORF7a (27736), ORF8 (27952), and NC region (30335, 31388, 32042, 32064, and 32093). The sample that has the most mutations are SARS-CoV-2 from Korea and Sweden with 8 mutation sites and there is a sample from USA Illinois that doesn't have a mutation. The highest mutation diversity occurred in a sample from Korea that mutates at ORF1a, ORF1b, S Protein, ORF7a, and ORF8. All other samples have at least 2 mutations sites and there is no identical mutation among samples (Fig. 1). Further studies that focused on mutation impact is required for revealing the implication of this mutation regarding its contribution to virulence, transmission, or therapy of SARS-CoV-2.

The result from the phylogenetic tree supplemented the alignment result. Generally, all of the analyzed sequences divided into four groups. The first group, which has the largest member consisted of SARS-CoV and SARS-CoV-2. The second group belongs to HcoV-OC43, Bovine CoV, HcoV-HKU1, and Rodent CoV. The third group was MERS CoV, while the fourth group was Canine CoV as out-group. Agreeing with the alignment result, sample number 3 of SARS-CoV-2 Jakarta was disjointed from sample numbers 1 and 2. Besides, SARS-CoV-2 from California also separated from Illinois, describing its different features so they were parted. Interestingly, Rodent CoV was placed in one clade with...
HCoV-HKU1, while Bovine CoV close with HCoV-OC43 in their branch (Fig. 2).

Figure 1. Alignment and mutation sites in SARS-CoV-2 from different countries compared to SARS-CoV, MERS-CoV and other Betacoronaviruses.
Figure 2. Phylogenetic tree of Betacoronavirus species and Canine Coronavirus as an out-group. Red number shows the bootstrap value that indicates its accuracy.

Figure 3. Haplotype network of 23 species of Betacoronavirus and an outgroup from Alphacoronavirus. Circles representing haplotypes, larger the circle more populations they have. Colour from circles shows the proportion of the populations from countries. Closer the distance between haplotypes the more genetic similarities.

Haplotype Network and Gene Flow

Haplotype reconstruction revealed that all analyzed Betacoronaviruses were divided into 3 main groups with total 11 haplotypes was discovered. This separation represents their subgenus, i.e. group 1 with SARS-CoV and SARS-CoV-2 from Sarbecovirus, group 2 with MERS-CoV from Merbecovirus, and group 3 with HCoV-HKU1, HCoV-OC43, Rodent Coronavirus and Bovine Coronavirus from Embecovirus. SARS-CoV-2 clumped in one haplotype, while SARS-CoV split into...
two different haplotypes with SARS-CoV from China has their one. In line with phylogenetic result, Rodent CoV and Bovine CoV have a close relationship with previously identified human coronavirus (Fig. 3). This was supported by sequence similarity results where Rodent and Bovine Coronavirus have relatively high similarity with HCoV-HKU1 (Fig. 4). Therefore, there is a big probability of future zoonosis from bovine and rodent due to adjacent similarity of their contained coronavirus. Gene flow analysis from DNASP shows that among SARS-CoV-2 used in this research, nucleotide diversity was 0.00022 and the average number of nucleotides was 6.5333. Gene flow strength between the China population and other populations was very high with Gst= 0.03498 and Nm= 6.90. Thus, we know that the transmission rate is high.

**Discussion**

Human-pathogenic coronavirus belongs to subfamily coronavirinae, which consists of Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus (Cui et al., 2019). With the current health crisis, Betacoronavirus has extensively studied due to its ability to cause pandemic over time. SARS-CoV, MERS-CoV, and SARS-CoV-2 are the member of Betacoronavirus which are familiar after causing global outbreak such as SARS, MERS, and Covid-19 (Cui et al., 2019). Genetic structure of Betacoronavirus, in particular SARS-CoV-2, has been studied previously (Chen et al., 2020; Khailany et al., 2020). These studies confirm that each of Betacoronavirus members have similarities among each other’s (Fig. 1). Furthermore, several point mutations in SARS-CoV-2 also discovered locating in ORF1a, ORF1b, S Protein, ORF6, ORF7a, ORF8, and NC region. Mutation in virus originated from genome replication error. In the case of DNA virus, there are proofreading that slows down replication error, but SARS-CoV-2 is an RNA virus that lack of proofreading (Peck & Lauring, 2018). Besides, the high rate of gene flow also influences virus mutation (Rubio et al., 2013). ORF1a encodes polypeptide pp1a, responsible to form the replication-transcription complex after processed by Main Protease (MPro) (Chen et al., 2020). Similarly, protein from ORF8 also contributes to the weakening of immune response by suppressing Major Histocompatibility Complex class I (MHC I) which has an important role for elimination of virus-infected cells (Zhang et al., 2020). Moreover, ORF8 protein showed a high possibility as a factor during inter-species transmission (Lau et al., 2015). Regarding those important roles of ORF1a, and ORF8, a mutation in those regions should be studied more to observe potential role in the severity and dispersion of coronavirus-related diseases.

**Figure 4.** BLAST result. (A) Rodent coronavirus with HCoV-HKU1, and (B) Bovine coronavirus with HCoV-OC43.

Either phylogenetic or haplotype network analysis consistently showed that the Betacoronavirus grouped into their respective subgenus (Fig. 2 and 3). Betacoronavirus is divided into 4 lineages, including lineage A (Embecovirus), B (Sarbecovirus), C (Merbecovirus), and D (Nobecovirus) (Drexler et al., 2014; Fahmi et al., 2020; Wassenaar & Zou, 2020). Embecovirus has two PL1 (papain-like protease), that are PL1PRO and PL2PRO, while the other Betacoronavirus only has 1 type of PLpro. It is known that Embecovirus also has Haemagglutinin Esterase (HE) gene (Woo et al., 2010). Merbecovirus and Sarbecovirus probably use DPP4 and ACE2 receptors, but in affinity, gradient binds to different host receptors (Wong et al., 2019).

Based on Figure 2-4, it showed that there are proximity between HCoV-OC43 with Bovine Coronavirus and HCoV-HKU1 with Rodent Coronavirus. Factors that increase the occurrence of zoonosis are phylogenetic closeness, RNA building receptor similarity, and interactions between human and animals, thus increasing a virus movement to a new host and become new virus strain (Olival et al., 2017). Thus, the relation among those species could indicate the possibility of future zoonosis to human. HCoV-OC43 and HCoV-HKU1 are viruses that infect human. HCoV-OC43 is a zoonotic virus from rodent that is infecting through bovine, while HCoV-HKU1 caused by a rodent (Cui et al., 2019; Ye et al., 2020). Bovine CoV is closely related to HCoV-OC43, so there is a possibility of zoonosis disease to human because of their genetic similarity up to 97.42% (Szczepanski et al., 2019). Bovine CoV has the same receptor as HCoV-OC43, so it is possible to transmit into human like HCoV-OC43 (Szczepanski et al., 2019). HCoV-OC43 is transmitted from rodent to human through bovine using a 9-O-acetic acid receptor on ciliated epithelial cells. The virus can adapt to its new host and began to mutate (Pyrz et al., 2010; Ye et al., 2020). Rodent CoV also has a possibility to start animal-human transmission due to the closeness to HCoV-HKU1 and has genetic similarity up to 82.90% (Fig. 4). That zoonosis can occur directly or indirectly through intermediate hosts to human (Ye et al., 2020).
The studies conclude that phylogenetic and haplotype network showed that beta coronavirus forms 4 groups of their origin from animal and human with high similarity of their genome. With the close relationship of Bovine CoV and Rodent CoV, it’s highly predicted that this virus has a probability to transmit into human in the future.

Acknowledgment

Thanks to Bioinformatics lecturer and assistant, especially Feri Eko Hernanto S.Si., M.Si for assisting of the in silico analyses and English writing.

References


Fahmi, M., Kubota, Y., & Ito, M. (2020). Nonstructural proteins N57b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases, 81, 104272. doi:10.1016/j.meegid.2020.104272


Rambaut, A. (2018). *Fgtree*, a graphical viewer of phylogenetic trees (Version 1.4.4) [Computer software].


