Sequence Analysis and Variations in the 5’ UTR genomic regions of Coronaviruses

Mohammed Bassyouni M. EL-Mahdi
Laboratory of Molecular Genetics and Molecular Biology, Zoology Department
Faculty of Science, South Valley University, Qena 83523, Egypt
*E.mail: melmahdi@svu.edu.eg

ARTICLE INFO
Article History
Received:25/10/2020
Accepted:25/12/2020
----------------------

Keywords:
Orthocoronavirinae, Coronaviruses, SARS CoV-2, 5’ UTR regions, molecular sequence variations

ABSTRACT
The 5’UTR genomic regions of coronaviruses were analysed regarding sequence variations, nucleotide constituents, and patterns of evolutionary course. The 5’ UTR of the viral genome highly conserved that appeared clearly between the closely related viruses. It possessed variable sites of varied nucleotides.

The 5’UTRs revealed long highly conserved regions (nt61 to nt107, nt211-nt236, nt262 to nt277). The longest one (nt61 to nt107) possessed nearly half of the core leader sequence (nt 61 to nt93) confirming its existence in viral genomes. The nucleotides composition exhibited favouritism towards AT(U) contents against GC contents which potentially is necessary for viral origin and replication activity. The length of 5′ UTR varies in studied cronaviruses, ranging from 245 bp (SARS coronavirus BJ01, AY278488.2) to 283 bp (hCoV-19_Chile, gaisd-EPI_ISL_445295). Additionally, the 5’ UTR region of Chile, EPI_ISL_445295 isolate included a foremost 18 nucleotides, these are not found in other isolates.

Phylogenetic analysis of 5’ UTR regions using the maximum likelihood method confirmed the close evolutionary distance and origin of betacronoviruses analysed. The most genetic closely isolates to the SARS-CoV-2 Wuhan-Hu-1 is the Bat coronavirus RaTG13. The SARS coronaviruses BJ01 and GD01 displayed the most distant betacronoviruses to SARS-CoV-2 that confirm their early arisen lineage. Finally, phylogenetic clustering in 5’ UTR regions for studied betacronoviruses revealed the genetic diversity of betacronoviruses and their high tendency towards frequent genetic mutations and gene recombination. This potentially leads to increased risk of interspecies transmission with viral evolution and accumulation of mutations.

INTRODUCTION
Globally, severe acute coronary syndrome 2 (SARS CoV-2) has become the world’s concern with increased risk of morbidity and mortality. The SARS CoV-2 is able to mutate rapidly, alter tissue tropism, cross the species barrier due to its ability to adapt to different epidemiological settings (Decaro et al., 2010; Wu et al., 2020; Huang et al., 2020; Zhou, et al., 2020). Coronaviruses (CoVs) are enveloped viruses with a single positive-stranded RNA genome (~26–32 kb in length) belonging to the subfamily Orthocoronavirinae (family: Coroniridae) (Weiss et al., 2005; Lai et al., 2007) including four genera: Alpha, Beta, Gamma and Deltacoronaviruses. Normally, the viral genome encodes four structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), several non-structural proteins and multiple unique accessory proteins (Woo et al., 2009; Perlman et al., 2009).
In December 2019, SARS-CoV-2 emerged in Wuhan City, China, causing severe respiratory illness and mortality (Chan et al., 2020). Early studies reported its potential transmission from bats as it has a high sequence identity of 96.3% with the bat coronavirus RaTG13 (Zhou et al., 2020).

SARS-CoV-2 is 29,903 nucleotides in length and making it the second-largest known RNA genome. The virus genome consists of two untranslated regions (UTRs) at the 5’ and 3’ ends and 11 open reading frames (ORFs) that encode 27 proteins (Wu et al., 2020; Chan et al., 2020; Ceraolo et al., 2020).

The 5’ and 3’ UTR regions of SARS-CoV-2 are comprised of 265 and 229 nucleotides, respectively. Both regions are necessitated for coronavirus infection, replication, transcription and functionally important for RNA-RNA interactions and binding of viral and cellular proteins (Yang et al., 2015). The leader sequence of 50-100 nucleotides at the 5’ end of the genome with an adjacent transcription regulatory sequence signify a unique feature of coronaviruses (Fields et al., 2001; Yang et al., 2015). The molecular sequence analysis and variations of 5’ UTR regions in SARS-CoV-2 may provide important clues about virus mutations, transmission, environmental adaptation and its pattern of global disease spread to avoid further escalation. Therefore, this study aimed to analyse the sequence variations, nucleotide constituents, and patterns of evolutionary course in the 5’ UTR genomic regions of Coronaviruses from various continentals’ isolates.

MATERAILS AND METHODS

Genomic Sequences Collection:

To obtain the 5’ UTR regions, available sequences from the GenBank database of the NCBI and GISAID database (https://gisaid.org) were searched for coronavirus genomes. Genomic sequences were considered complete if they were polyadenylated at their 3’ ends. In total, twelve (12) reference genomes sequences of Alpha and Betacoronavirus isolates/strains were selected (Table 1) and downloaded.

Table 1. Accession numbers and identifier of coronavirus genomic sequences used to retrieve the 5’ UTR regions in this study.

<table>
<thead>
<tr>
<th>Accession Numbers &amp; identifier of Coronaviruses</th>
<th>Subfamily/genera</th>
<th>5’ UTR Length</th>
<th>Genome Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPI ISL 445295.hCoV-19 Chile/South America</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>283 bp</td>
<td>29.900</td>
</tr>
<tr>
<td>EPI ISL 515182.hCoV-19 Ghana/Africa</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>261 bp</td>
<td>29.899</td>
</tr>
<tr>
<td>EPI ISL 444998.hCoV-19 Guam/Oceania</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>265 bp</td>
<td>29.903</td>
</tr>
<tr>
<td>EPI ISL 412974.hCoV-19 Italy/Europe</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>265 bp</td>
<td>29.903</td>
</tr>
<tr>
<td>AY278488.2 SARS coronavirus BJ01</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>245 bp</td>
<td>29.725</td>
</tr>
<tr>
<td>MN999032.1 Bat coronavirus RaTG13</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>250 bp</td>
<td>29.855</td>
</tr>
<tr>
<td>AY278489.2 SARS coronavirus GD01</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>248 bp</td>
<td>29.757</td>
</tr>
<tr>
<td>MG772933.1 Bat-SL-CoVZC45</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>264 bp</td>
<td>29.802</td>
</tr>
<tr>
<td>MG772934.1 Bat-SL-CoVZC21</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>264 bp</td>
<td>29.732</td>
</tr>
<tr>
<td>MT365419.1 SARS-CoV-2 Human/USA/CA-CZB</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>264 bp</td>
<td>29.902</td>
</tr>
<tr>
<td>NC_045512.2 SARS coronavirus2/Wuhan-Hu-1</td>
<td>Orthocoronavirinae, Betacoronavirus</td>
<td>265 bp</td>
<td>29.903</td>
</tr>
<tr>
<td>NC_002645.1 RHCoV-29E</td>
<td>Orthocoronavirinae, Alphacoronavirus</td>
<td>265 bp</td>
<td>27.317</td>
</tr>
</tbody>
</table>

Data Analysis:

To analyse the 5UTR region, nucleotide sequence upstream to initiation codon (ATG) were retrieved from each reference genomic sequence. The retrieved 5 UTRs regions were edited and viewed using Bioedit (Hall, 1991) and SnapGene Viewer v3.2.1 (Freeware License, SnapGene® software/GSL Biotech) softwares under defaults options. Sequence multiple
alignments were conducted and analysed using the Mega6 software (Tamura et al., 2013) and included the DNA sequence variances and nucleotide composition.

To investigate the genomic relation and evolutionary history among retrieved 5’ UTR sequences, phylogenetic tree reconstruction was carried out using the maximum likelihood (Tamura et al., 2004) clustering method and rooted with the alphacoronavirus (NC_002645.1 RHCoV-229E) as an outgroup. Trees were constructed with node/branch support of 1000 bootstrap replicates (Felsenstein, 1985) where the branch length measured the number of substitutions per site. The best model for nucleotide substitution rate for sequences was selected by Maximum Likelihood fits of 24 different nucleotide substitution models (Nei and Kumar, 2000). The selection was based on Bayesian information criterion (BIC) and Akaike Information Criterion (AIC).

RESULTS

Sequence Alignment of 5’ UTR Regions:

Multiple sequence alignment performed on eleven (11) retrieved 5’ UTR regions (Fig. 1) displayed a distinct conserved region up to 81% (boxed), those were clearly conserved among studied virus isolates. Also, there were variable sequence sites of 19% (coloured). Data shows that the majority of conserved stretches within the evaluated 5’UTRs were concentrated after the first 25 base pairs. There are long highly conserved regions within the whole 5’ UTR in the analysed coronavirus isolates. These are from nucleotide 61 to nucleotide 107 (61nt-107nt), from nucleotide 211 to nucleotide 236 (211nt-236nt) and from nucleotide 262 to nucleotide 277 (nt262 to nt277). Comparably, the longest one (nt61 to nt107) possess nearly half (nt 61 to nt 93) of the core leader sequence reported (Li et al., 2005; Baldassarre et al., 2020) which mainly conserved among studied strains. Sequence alignments showed that the sequence region from nt1 to nt18 is only present in the 5’ UTR region belonged to the gaisd-EPI_ISL_445295 isolate (Chile, Punta Arenas, South America) and not in any of other analysed isolates.

![Fig. 1:Sequences alignments in the 11 5’ UTR coronaviruses genomic regions.](image-url)

The conservation pattern among the represented 5’ UTR genomic sequences are boxed, while the variable sites are coloured. The leader sequence is shown underlined, spanning sequence area from nucleotide (nt) nt 32 to nt 93.
Nucleotide Composition of 5' UTR Region in Coronavirus:

Alignments of eleven (11) of 5' UTR genomic regions upstream to initiation codon (ATG) produced an average nucleotide length of 261.3 base pairs and a consensus length of 285 sites (Fig. 1) which included base pairs, gaps and indel (insertion/deletion) sites. In average, nucleotides composition (Table 2) was T(U) =30.3, C=23.8, A=25.0 and G=20.8. The G+C=44.66% and A+T=55.34% revealed a nucleotide favouritism towards AT contents.

<p>| DNA Nucleotide constitutions of 5' UTR region analyzed for 11 Coronavirus strains. |
| C= conserved; V= variable; PI= parsimony informative; S= singleton sites |</p>
<table>
<thead>
<tr>
<th><strong>Corona virus Strains</strong></th>
<th><strong>Nucleotide constitutions</strong></th>
<th><strong>T(U)</strong></th>
<th><strong>C</strong></th>
<th><strong>A</strong></th>
<th><strong>G</strong></th>
<th><strong>Total</strong></th>
<th><strong>G+C%</strong></th>
<th><strong>A+T%</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>EPI ISL 445295.1 hCoV-19 Chile/South America</td>
<td>31.1</td>
<td>23.7</td>
<td>24.4</td>
<td>20.8</td>
<td>283.0</td>
<td>44.52</td>
<td>55.48</td>
<td></td>
</tr>
<tr>
<td>EPI ISL 515182.1 hCoV-19 Ghana/Africa</td>
<td>29.9</td>
<td>24.1</td>
<td>24.9</td>
<td>21.1</td>
<td>265.0</td>
<td>45.21</td>
<td>54.79</td>
<td></td>
</tr>
<tr>
<td>EPI ISL 444998.1 hCoV-19 Guam/Oceania</td>
<td>30.2</td>
<td>23.8</td>
<td>25.3</td>
<td>20.8</td>
<td>265.0</td>
<td>45.43</td>
<td>54.57</td>
<td></td>
</tr>
<tr>
<td>EPI ISL 412974.1 hCoV-19 Italy/Europe</td>
<td>30.2</td>
<td>23.8</td>
<td>25.3</td>
<td>20.8</td>
<td>265.0</td>
<td>45.43</td>
<td>54.57</td>
<td></td>
</tr>
<tr>
<td>EPI ISL 478488.2 SARS coronavirus J01</td>
<td>27.8</td>
<td>24.9</td>
<td>26.1</td>
<td>21.2</td>
<td>245.0</td>
<td>46.12</td>
<td>53.88</td>
<td></td>
</tr>
<tr>
<td>MV936532.1 Bat coronavirus RatG13</td>
<td>30.0</td>
<td>25.0</td>
<td>22.5</td>
<td>25.0</td>
<td>200.0</td>
<td>44.80</td>
<td>55.20</td>
<td></td>
</tr>
<tr>
<td>EPI ISL 478489.2 SARS coronavirus GD01</td>
<td>27.8</td>
<td>25.0</td>
<td>22.5</td>
<td>25.0</td>
<td>200.0</td>
<td>44.80</td>
<td>55.20</td>
<td></td>
</tr>
<tr>
<td>MG772933.1 Bat-SL-CoVZC45</td>
<td>33.0</td>
<td>22.7</td>
<td>23.9</td>
<td>20.5</td>
<td>265.0</td>
<td>43.18</td>
<td>56.82</td>
<td></td>
</tr>
<tr>
<td>MG772933.1 Bat-SL-CoVZC21</td>
<td>33.0</td>
<td>22.7</td>
<td>23.9</td>
<td>20.5</td>
<td>265.0</td>
<td>43.18</td>
<td>56.82</td>
<td></td>
</tr>
<tr>
<td>MT581419.1 SARS-CoV-2 Human/USA/CAS-CZB</td>
<td>30.3</td>
<td>23.9</td>
<td>25.0</td>
<td>20.8</td>
<td>265.0</td>
<td>44.70</td>
<td>55.30</td>
<td></td>
</tr>
<tr>
<td>NC-045512.2 SARS-CoV-2 Wuhan-Hu-1</td>
<td>30.3</td>
<td>23.2</td>
<td>25.3</td>
<td>20.8</td>
<td>261.3</td>
<td>44.66%</td>
<td>55.34</td>
<td></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>30.3</td>
<td>23.8</td>
<td>25.0</td>
<td>20.8</td>
<td>261.3</td>
<td>44.66%</td>
<td>55.34</td>
</tr>
</tbody>
</table>

Molecular Phylogenetic Analysis of Coronavirus 5' UTR Regions:

Phylogenetic analysis of eleven (11) 5' UTR genomic sequences produced ML tree that rooted with the alphacronavirus NC_002645.1 RHCoV-229E (Fig. 2). The suitable nucleotide substitution model was found to be the K2 (Kimura 2-parameter) model (BIC=-741.725; transition/transversion bias (R)=1.79; Nucleotide frequencies of 0.250 for f(A), f(T), f(C), and f(G). The pairwise genetic distances among the 5’ UTR regions (11 analysed betacoronaviruses isolates + the NC_002645.1 RHCoV-229E as out group) which computed by the K2 (Kimura 2-parameter) model is shown in table 3. Data showed that the closely related isolates have less distance values comparing to the distantly related ones as shown in the table. The SARS-CoV-2Wuhan-Hu-1(NC-045512.2), SARS-CoV-2 Human/USA (MT385419.1), hCoV-19_Italy/Europe (EPI ISL 412974), hCoV-19_Guam (EPI ISL 444998), hCoV-19_Ghana (EPI ISL 515182) and hCoV-19_Chile (EPI ISL 445295) possessed less distance values ranged from 0.000 to 0.100 among them. While a higher distance value of
(0.824) recorded for the SARS coronavirus BJ01 (AY278488.2) and SARS coronavirus GD01 (AY278489.2) with the alphacoronavirus NC_002645.1_RHCoV-229E.

The ML phylogenetics tree (Fig. 2) with the highest log likelihood (-741.7249), demonstrated three groups. All betacoronavirus isolates were closely assembled in groups A and B. Group A included SARS coronavirus BJ01 (AY278488.2) and SARS coronavirus GD01 (AY278489.2). Group B contained two subgroups; Bat SLCoVZXC21 (MG772934.1) and Bat SLCoVZC45 (MG772933.1) which closely assembled (subgroup SGB1) showing higher genetic relations. The subgroup B2 (SGB2) are closely clustered into two evolutionary lines, SGB2-1 and SGB2-2. The SGB2-1 contained the Bat coronavirus RaTG13 (MN996532.1) while, the SGB2-2 comprised the SARS-CoV_2Wuhan-Hu-1(NC-045512.2), SARS-CoV-2 Human/USA (MT385419.1), hCoV-19_Ireland/Europe (EPI_ISL_412974), hCoV-19_Guam (EPI_ISL_444998), hCoV-19_Ghana (EPI_ISL_515182) and hCoV-19_Chile (EPI_ISL_445295) which reflected their close phylogeny relation and genetic evolutionary origin.

The closest located isolate to subgroup SGB2-1 that owned the Bat coronavirus RaTG13 is the SARS-CoV_2Wuhan-Hu-1. Both SARS coronavirus BJ01 (AY278488.2) and SARS coronavirus GD01 (AY278489.2) (group A) displayed a distant lineage to members of group B, but they share a common evolutionary lineage with them. Additionally, phylogenetic analysis separately placed the alphacoronavirus RHCoV-229E (NC_002645.1) (group C) which displayed its distant evolutionary relationships to other analysed coronaviruses.

Table 3. The pairwise genetic distance involving 12 5UTR genomic regions (11 analysed Coronaviruses + the NC_002645.1 RHCoV-229E as an outgroup). Analyses were conducted using the maximum composite likelihood based on the K2 (Kimura 2-parameter) model.
Fig. 2: Molecular phylogenetic analysis by maximum likelihood method using 5’ UTR genomic regions for 11 selected coronaviruses and NC_002645.1 RHCoV-229E as an outgroup, conducted using the K2 (Kimura 2-parameter) model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The bootstrap support of 1000 replicates are shown next to the branches. Evolutionary analyses were conducted in MEGA6. SGB1: Subgroup B1; SGB2: Subgroup B2; SGB2-1: Subgroup B2-1; SGB2-2: Subgroup B2-2.

DISCUSSION

Herein, the 5’ UTR genomic regions of coronavirus from various locations were investigated. Analysis included the sequence variations, nucleotide constituents, and patterns of evolutionary course.

Sequence conservation of 5’ UTR regions among studied viral isolates shown to be near 81% (boxed) which was maintained in the majority of them. Whoever, isolates exhibited nucleotide variations of variable sites (coloured in the alignments) which are significantly different in length and nucleotide sequences. Such pattern is commonly observed among the same closely related isolates (e.g SARS coronavirus BJ01 (AY278488.2) and SARS coronavirus GD01 (AY278489.2); also, between the Bat SLCoVZXC21 (MG772934.1) and Bat SLCoVZC45 (MG772933.1). As reported, these conserved stretches may harbour varieties of control elements which points to genetic interaction and involvements for virus viability, replication, translational efficiencies, and viral rapid mutation rate (de Haan et al., 2002; Goebel et al., 2007; Zust et al., 2008; de Groot et al., 2012). The conserved genomic segments permit the constancy between conserved and variable viral elements, which potentially necessary for promoting escape from immune selection and rapid adaptation to novel environments (Domingo et al., 1997; Kaur et al., 2020; Chan et al., 2020). Additionally, the occurrence of nucleotide variations (e.g variable sites) in the 5’UTR regions analysed, potentially due to own viral means of adaptation for local genomic replication, indicating a rapid transmission and better adaptation of SARS-CoV-2 (COV-19) continently (Domingo et al., 1997; Chan et al., 2020).

Data clearly demonstrated differences in length and sequence of 5’ UTR regions among analysed coronavirus isolates, varying from 245 bp (SARS coronavirus BJ01,
AY278488.2) to 283 bp (hCoV-19_Chile, gaisd-EPI_ISL_445295). This finding is consistent with the 5′ UTR variable lengths in earlier reported viral genomes (Yang et al., 2015; Madhugiriet al., 2018; Chan et al., 2020; Wu et al., 2020; Chan et al., 2020).

The longest continuous conserved area was from nt61 to nt107, nt211-nt236, and nt262 to nt277. These stretches are comparably long highly conserved regions within the whole 5′ UTRs analysed. The longest one (nt61 to nt107) nearly possess half of the core leader sequence (nt 61 to nt93) which confirmed the presence of the leader sequence in the viral 5′ UTR regions that is consistent with other studies (Zeng et al., 2003; Li et al., 2005; Sola et al., 2015; Baldassarreet al., 2020). Conspiracy of the leader sequence with transcription regulatory sequence (TRS) consider a distinctive constituent of coronaviruses (Marraet al., 2003; Rota et al., 2003) and is critically necessitated for gene expression during viral replication (Sawickiet al., 1998; Lai et al., 1997). Taking into consideration, that such model of conservation reaches 100% between SARS-CoV_2Wuhan-Hu-1 and SARS-CoV-2 Human/USA (MT385419.1) isolates.

Our result demonstrated that the 5′ UTR region of Chile, EPI_ISL_445295 has a foremost 18 nucleotides which were found to be missing in all other analysed isolates. Potentially, this may a possible regional virus self-attitude as a requirement for its local viral environmental adaptations in the infected host that linked to the host’s replication machinery (Zhang et al., 2006). Additionally, the missing nucleotides from the 5′ UTR regions are reported, where nucleotide deletion and substitution could be common (Sharif and Dey, 2020).

Nucleotide composition of studied 5′ UTR region exhibited nucleotides favouritism towards AT(U) contents compared to GC contents. Potentially, the observations of AT favouritism reflected that such AT regions possibly encode for well-defined, conserved RNA structures which are necessary for establishing molecular machinery and formation of long-range interactions. Studies reported the importance of AT-rich regions as they are critical for the origin replication activity (Ranganet al., 2020; Li et al., 2008; Yang et al., 2006).

Genetically, most closely related viruses harbour similar genetic makeup. Here phylogenetic tree constructed using the maximum likelihood method intimately assembled the studied betacoronaviruses in two groups (A, B) which mirror their close evolutionary distance and origin. Among them, the most genetic closely isolates are the Bat coronavirus RaTG13 and SARS-CoV_2Wuhan-Hu-1. A study reported that the most closely related virus to SARS-CoV-2 is the RaTG13 (Zhou et al., 2020). Data confirmed that SARS-CoV-2 (COV-19) isolates from different continents belong to the betacoronavirus and showed closer clustering to the bat-SL-CoVZC45 and bat-SL-CoVZXC21 viruses which reflected their genetic resemblances and evolutionary relation. As reported, the SARS-CoV-2 forms a distinct lineage with Bat-SARS-like coronaviruses that belong to order Nidovirales sharing high level of nucleotide similarities (Zhou et al., 2020; Chan, et al., 2020; Paraskeviset al., 2020). Additionally, studies concluded the identification of bats as reservoir hosts of SARS-CoV-2. Hao et al., (2020) and confirm the zoonotic origin of the virus.

The phylogenetic tree displayed the distant assembling of SARS coronavirus BJ01 and SARS coronavirus GD01 relative to COV-19 isolates. Potentially, they may represent the most distant betacoronavirus lineage to COV-19 confirming their early arisen as reported (Qin et al., 2003a; Qin et al., 2003b; Shengli et al., 2003). These two betacoronaviruses may genetically close to members of group B with common evolutionary lineage and similar genetic constituents.

Finally, phylogenetic clustering in 5′ UTR regions of studied betacoronaviruses revealed their genetic diversity and high tendency towards frequent genetic mutations. This potentially led to an increased risk of interspecies transmission with viral evolution.
and accumulation of mutations.

**Conclusion**

The 5′ UTR of the studied betacoronaviruses is highly conserved that appeared clearly between the closely related viruses, and possessed nucleotide variable sites with different length and nucleotide composition. The 5′UTRs revealed long highly conserved regions (nt61 to nt107, nt211- nt236, nt262 to nt277). The longest one (nt61 to nt107) posse nearly half of the core leader sequence (nt 61 to nt93) confirming the existence of the leader sequence in all viral 5′ UTR regions analysed. The nucleotides composition exhibited preference towards AT(U) contents against GC contents which potentially is needed for the viral origin and replication activity. The length of 5′ UTR varies in analysed betacoronaviruses, ranging from 245 bp (SARS coronavirus BJ01, AY278488.2) to 283 bp (hCoV-19_Chile, gaisd-EPI_ISL_445295). Additionally, the 5′ UTR region of Chile, EPI_ISL_445295 isolate included a foremost 18 nucleotides, these are found to be missing in other isolates.

Phylogenetic analysis using the maximum likelihood method confirmed the close evolutionary distance and origin of Betacoronovirus analysed. The closest genetic isolates to the SARS-CoV-2Wuhan-Hu-1 is the Bat coronavirus RaTG13. The SARS coronavirus BJ01 and SARS betacoronaviruses GD01 are distantly betacoronavirus lineage to COV-19 reflecting their early arisen. Finally, phylogenetic clustering revealed the genetic diversity of betacoronovirus and their high tendency towards frequent genetic mutations and gene recombination. This may cause an increased risk of interspecies transmission with viral synchronized evolution and accumulation of mutations.

**Acknowledgment**

This work was funded by the South Valley University (research support), Qena, EGYPT. Author is thankful to Professor A.S.M Mustfa, Zoology Department, Faculty of Science, South Valley University, Egypt for his valuable comments.

**REFERENCES**


Mohammed Bassyouni M. EL-Mahdi


Zhang, J.J., Huang, A.L., Shi, X.L., Zhang,


**ARABIC SUMMARY**

التعابين النوكليوتيدى و التنوع في المناطق الجينومية (5’UTRs) للفيروسات التاجية (betacronoviruses)

محمد بسيوني محمد المهدى
معمل الوراثة الجزيئية وبيولوجيا الجزيئيات - قسم علم الحيوان - كلية العلوم
جامعة جنوب الوادى - جمهورية مصر العربية

في هذه الدراسة تم تحليل المناطق الجينومية (5’UTRs) للفيروسات التاجية فيما يتعلق باختلافات التسلسل النوكليوتيدى ووكانتها وكذلك أداء العلامة التطورية الوراثية. أوضح الدراسة أن تلك المناطق (5’UTRs) محفوظة بدرجة كبيرة داخل المادة الوراثية الفيروسية والواضح بين الفيروسات وثيقة الصلة، كما أشارت لاحتواء تلك المناطق لمواقع متغيرة من التتابع النوكليوتيدى.

أظهرت الدراسة وجود ثلاث سترات محفوظة ذات تتابع نوكليوتيدى طويل في تلك المنطقة كالاتى: من nt61 إلى nt107 (ذات التتابع النوكليوتيدى ذات الطول متبقي من نصف التسلسل leader sequence (61nt) إلى 107nt) و أيضاً من nt262 إلى nt277 (ذات التتابع النوكليوتيدى ذات الطول متبقي من نصف التسلسل leader sequence (61nt) إلى 93nt) للفيروس التاجي SARS-CoV-2 Wuhan-Hu-1، والذي ربما يكون أساسيًا للنشاط الفيروسى و نسخ المتماثل.

بينت الدراسة اختلافات في حجم مناطق (5’UTRs) للفيروسات التاجية المروعة (hCoV-19_Chile) الذي تتراوح من 245 زوج قاعدة للفيروس SAR-CoV-2 إلى 238 زوج قاعدة للفيروس BJ01، والذي أدى تحتوي على 18 زوج قاعدة أضافية مقارن بالفيروسات الأخرى.

أكد التحليل الوراثي باستخدام طريقة الاحتمال القصوى للكتابة استخدام طريقة الاحتمال القصوى للكتابة لتفاوت النوكليوتيدى و (5’UTRs) من المناطق الجينومية (5’UTRs) الفيروسات التاجية (betacronoviruses) حيث كان للفيروس التاجي SARS-CoV-2 و RaTG13 هو فيروس الخفاشات الناجى التي ربما بدأ تطورها في الفيروس السارس (BJ01 and GD01) من الفيروس التاجي SARS-CoV-CoV-2. كما أظهرت كل من الوراثي و حدوث التغيرات الجينية و الاستنساخ الفيروسى المتجدد أن التغير و التقويض يمكن أن يؤدي إلى زيادة مخاطر انتقال العدوى بين الأنواع و المجتمعات.