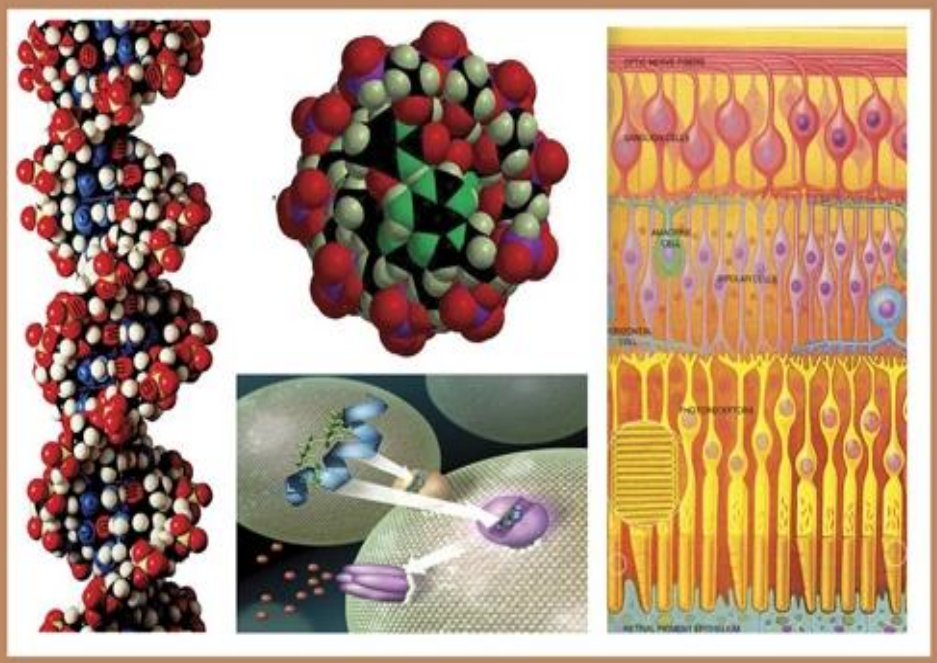
EGYPTIAN ACADEMIC JOURNAL OF

BIOLOGICAL SCIENCES

C

PHYSIOLOGY & MOLECULARBIOLOGY



ISSN
2090-0767

WWW.EAJBS.EG.NET

Vol. 12 No. 2 (2020)

Citation: *Egypt.Acad.J.Biolog.Sci. (C.Physiology and Molecular biology) Vol. 12(2) pp201-211(2020)*

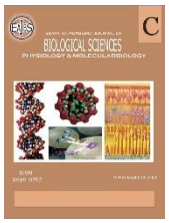
Egypt. Acad. J. Biolog. Sci., 12(2):201-211(2020)

Egyptian Academic Journal of Biological Sciences

C. Physiology & Molecular Biology

ISSN 2090-0767

www.eajbsc.journals.ekb.eg



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: Coronaviridae) (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.

Accession Numbers & identifier of Coronaviruses	Subfamily/genera	5' UTR Length	Genome Length
EPI ISL 445295/hCoV-19 Chile/South America	Orthocoronavirinae, Betacoronavirus	283 bp	29.900
EPI ISL 515182/hCoV-19 Ghana/Africa	Orthocoronavirinae, Betacoronavirus	261 bp	29.899
EPI ISL 444998/hCoV-19 Guam/Ocenia	Orthocoronavirinae, Betacoronavirus	265 bp	29.903
EPI ISL 412974/hCoV-19 Italy/Europe	Orthocoronavirinae, Betacoronavirus	265 bp	29.903
AY278488.2 SARS coronavirus BJ01	Orthocoronavirinae, Betacoronavirus	245 bp	29.725
MN996532.1 Bat coronavirus RaTG13	Orthocoronavirinae, Betacoronavirus	250 bp	29.855
AY278489.2 SARS coronavirus GD01	Orthocoronavirinae, Betacoronavirus	248 bp	29.757
MG772933.1 Bat-SL-CoVZC45	Orthocoronavirinae, Betacoronavirus	264 bp	29.802
MG772934.1 Bat SL-CoVZXC21	Orthocoronavirinae, Betacoronavirus	264 bp	29.732
MT385419.1 SARS-CoV-2/Human/USA/CA-CZB	Orthocoronavirinae, Betacoronavirus	264 bp	29.902
NC-045512.2 SARS coronavirus2Wuhan-Hu-1	Orthocoronavirinae, Betacoronavirus	265 bp	29.903
NC_002645.1_RHCoV-229E	Orthocoronavirinae,Alphacoronavirus	265 bp	27.317

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, phylogentic tree reconstruction was carried out using the maximum likelihood (Tamura et al., 2004) clustering method and rooted with the alphacoronvorus (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% (colured). Data shows that the majority of conserved stretches within the evaluated 5UTRs 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*etal.*, 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.

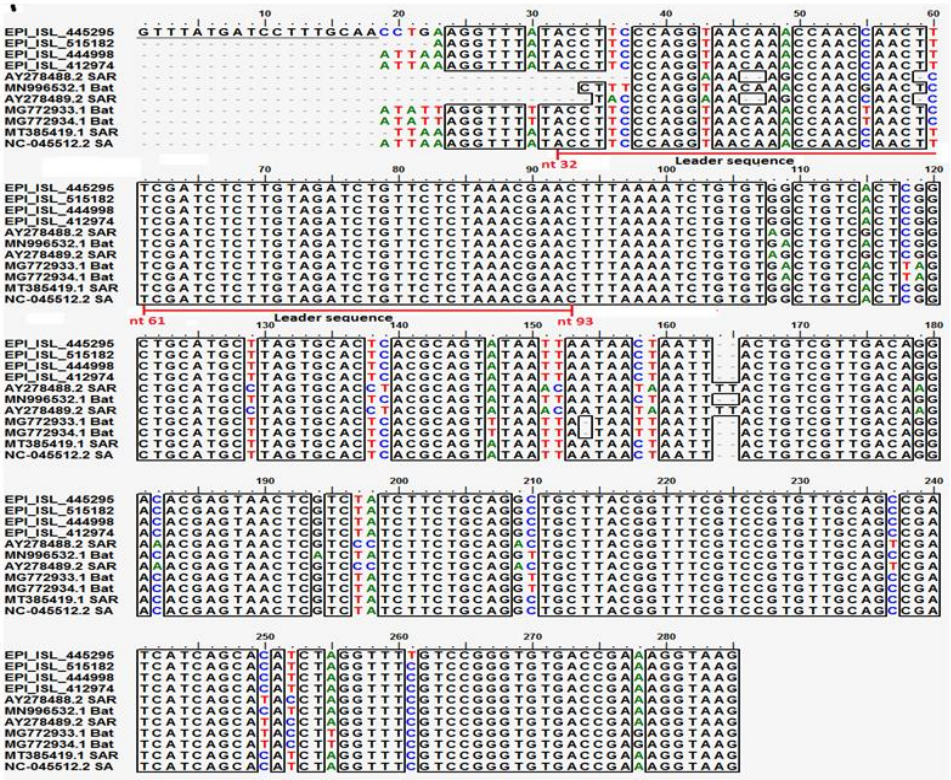


Fig. 1:Sequences alignments in the 11 5' UTR coronaviruses genomic regions. The conservation pattern among the represented 5' UTR genomic sequences are boxed, while the variable sites are coloured. The leader sequence is shown underlined, spanning sequence area from nucleotide (nt) nt 32 to nt 93.

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.

Analysis of 285 sites of coronavirus 5’ UTR region revealed 229 (80.35%) conserved sites and 38 (13.33%) variable nucleotides (coloured in the alignments). From the variable nucleotides, 32 (11.23%) were parsimony informative, and 6 (2.10%) were singletons. The great majority of the 5' UTR sites were conserved (80.35%), but there was a sequence variability of 13.33% was observed (Table 2). Nucleotide analysis also showed indels within 5' UTR regions in studied betacoronaviruses.

Table 2. DNA Nucleotide constitutions of 5 UTR region analyzed for 11 Corona virus strains. C= conserved; V= variable; PI= parsimony informative; S= singleton sites

Corona virus Strains/ Nucleotide constitutions	T(U)	C	A	G	Total	G+C%	A+T%	C	V	PI	S
EPI ISL 445295/hCoV-19 Chile/South America	31.1	23.7	24.4	20.8	283.0	44.52	55.48	229	38	32	6
EPI ISL 515182/hCoV-19 Ghana/Africa	29.9	24.1	24.9	21.1	261.0	45.21	54.79				
EPI ISL 444998/hCoV-19 Guam/Ocenia	30.2	23.8	25.3	20.8	265.0	44.53	55.47				
EPI ISL 412974/hCoV-19 Italy/Europe	30.2	23.8	25.3	20.8	265.0	44.53	55.47				
AY278488.2 SARS coronavirus BJ01	27.8	24.9	26.1	21.2	245.0	46.12	53.88				
MN996532.1 Bat coronavirus RaTG13	30.0	24.0	25.2	20.8	250.0	44.80	55.20				
AY278489.2 SARS coronavirus GD01	27.8	25.0	26.2	21.0	248.0	45.97	54.03				
MG772933.1 Bat-SL-CoVZC45	33.0	22.7	23.9	20.5	264.0	43.18	56.82				
MG772934.1 Bat SL-CoVZXC21	33.0	22.7	23.9	20.5	264.0	43.18	56.82				
MT385419.1 SARS-CoV-2/Human/USA/CA-CZB	30.3	23.9	25.0	20.8	264.0	44.70	55.30				
NC-045512.2 SARS coronavirus2Wuhan-Hu-1	30.2	23.8	25.3	20.8	265.0	44.53%	55.47				
Average	30.3	23.8	25.0	20.8	261.3	44.66%	55.34				

Molecular Phylogenetic Analysis of Coronavirus 5’ UTR Regions:

Phylogentic analysis of eleven (11) 5' UTR genomic sequences produced ML tree that rotted with the alphacrona virus NC_002645.1 RHCoV-229E (Fig. 2). The suitable nucleotide substitution model was found to be the K2 (Kimura 2-parameter) model (BIC=1657.563; AIC=1527.823; lnL= -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.

TheSARS-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 alphacoronvirus NC_002645.1_RHCoV-229E.

The ML phylogenetics tree (Fig. 2) with the highest log likelihood (-741.7249), demonstrated three groups. All betacronovirus 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_Italy/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 alphacronovirus 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.

Corona viruses Strains Accession Numbers		1	2	3	4	5	6	7	8	9	10	11	12
1	EPI_ISL_445295/hCoV-19_Chile/South_America	--											
2	EPI_ISL_515182/hCoV-19_Ghana/Africa	0.004	--										
3	EPI_ISL_444998/hCoV-19_Guam/Ocenia	0.004	0.000	--									
4	EPI_ISL_412974/hCoV-19_Italy/Europe	0.004	0.000	0.000	--								
5	AY278488.2_SARS_coronavirus_BJ01	0.100	0.095	0.095	0.095	--							
6	MN996532.1_Bat_coronavirus_RaTG13	0.027	0.022	0.022	0.022	0.110	--						
7	AY278489.2_SARS_coronavirus_GD01	0.100	0.095	0.095	0.095	0.000	0.110	--					
8	MG772933.1_Bat-SL-CoVZC45	0.050	0.046	0.046	0.046	0.104	0.036	0.104	--				
9	MG772934.1_Bat_SL-CoVZXC21	0.050	0.046	0.046	0.046	0.104	0.036	0.104	0.000	--			
10	MT385419.1_SARS-CoV-2/Human/USA/North_America	0.004	0.000	0.000	0.000	0.095	0.022	0.095	0.046	0.046	--		
11	NC-045512.2_SARS-CoV_2Wuhan-Hu-1/China/Asia	0.004	0.000	0.000	0.000	0.095	0.022	0.095	0.046	0.046	0.000	--	
12	NC_002645.1_RHCoV-229E	0.796	0.780	0.780	0.780	0.824	0.792	0.824	0.774	0.774	0.780	0.780	--

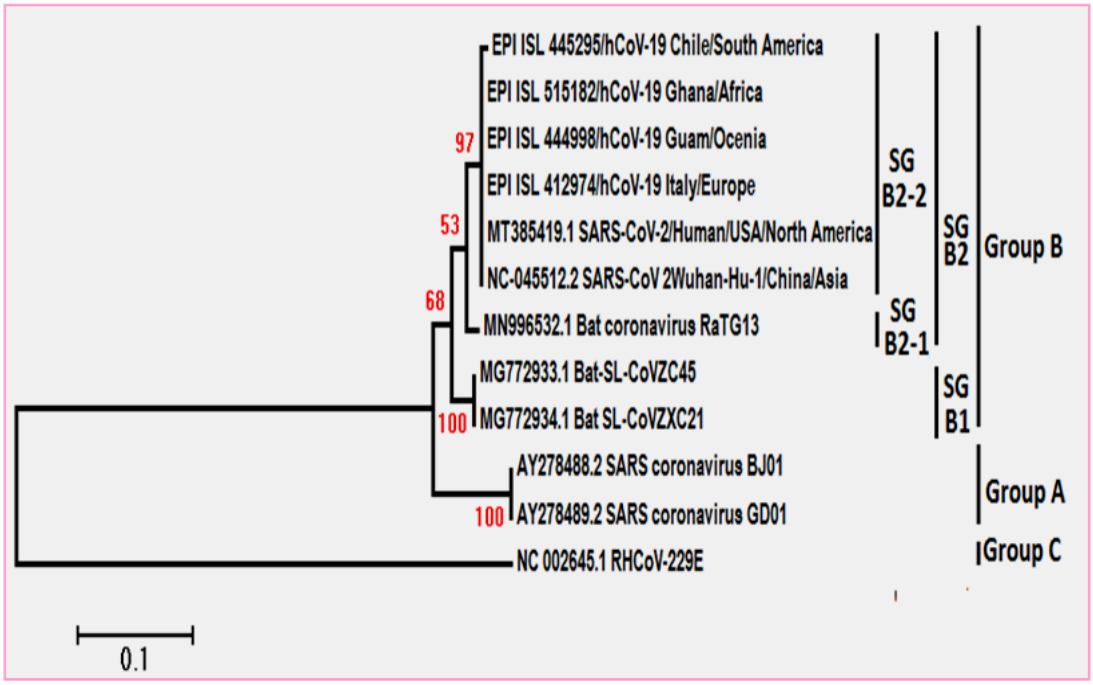


Fig. 2: Molecular phylogenetic analysis by maximum likelihood method using 5' UTR genomic regions for 11selected coronaviruses and NC_002645.1 RHCoV-229E as an outgroup, conducted using the K2 (Kimura 2-parameter) model. The tree is drawn to scale, with branch lengths measured in the number of substitutions 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; Madhugiri*et 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; Baldassarre*et. al.,* 2020). Conspiracy of the leader sequence with transcription regulatory sequence (TRS) consider a distinctive constituent of coronaviruses (Marra*et al.,* 2003; Rota *et al.,* 2003) and is critically necessitated for gene expression during viral replication (Sawicki *et 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 (Rangan*et 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 betacronoviruses 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 continentals 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; Paraskevis*et 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 phylogentic 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 betacronovirus linage 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 betacronoviruses 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 Betacronovirus 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 betacronovirus linage to COV-19 reflecting their early arisen. Finally, phylogenetic clustering revealed the genetic diversity of betacronovirus 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

Baldassarre, A., Paolini, A., Bruno, S. P., Felli, C., Tozzi, A. E., &Masotti, A. (2020): Potential use of noncoding RNAs and innovative therapeutic strategies to target the 5'UTR of SARS-CoV-2. *Epigenomics,* 12(15): 1349–1361. https://doi.org/10.2217/epi-2020-0162.

Bi, S., Qin E., Xu Z., Li W., Wang J., Hu Y., Liu Y., Duan S., Hu J., Han Y. et al. (2003): Complete genome sequences of the SARS-CoV: the BJ Group (Isolates BJ01-BJ04). *Genomics Proteomics Bioinformatics,* 1(3):180-92. doi: 10.1016/s1672-0229 (03) 01023-4.

Ceraolo, C., Giorgi, F.M. (2020): Genomic variance of the 2019-nCoV coronavirus. *Journal of medical virology,*92: 522–528.

Chan, A. P., Choi, Y., and Schork, N. J. (2020): Conserved Genomic Terminals of SARS-CoV-2 as Co-evolving Functional Elements and Potential Therapeutic Targets. *bioRxiv*, [preprint] https://doi.org/10.1101/2020.07.06.190207.

Chan, J.F., Kok, K.H., Zhu, Z., Chu, H., To, K.K., Yuan, S., Yuen, K.Y. (2020): Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections,*9: 221–236.

Chan, J.F., Yuan, S., Kok, K.H., To, K.K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C., Poon, R.W., et al. (2020): A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet,* 395, 514–523.

de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C.Y., Ziebuhr, J. (2012 Family Coronaviridae. In: King, A.M.Q.,

Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus taxonomy. *Elsevier, Amsterdam,* pp. 806–828.

de Haan, C.A., Volders, H., Koetzner, C.A., Masters, P.S., Rottier, P.J., (2002): Coronaviruses maintain viability despite dramatic rearrangements of the strictly conserved genome organization. *Journal of virology,* 76: 12491–12502.

Decaro, N., Mari, V.; Elia, G., Addie, D.D., Camero, M., Lucente, M.S., Martella, V., Buonavoglia, C. (2010): Recombinant canine coronaviruses in dogs, Europe. *Emerging Infectious Diseases,*16: 41–47.

Domingo, E., and J. J. Holland. (1997): RNA virus mutations and fitness for survival. *Annual Review of Microbiology,* 51:151-178.

Felsenstein, J., (1985): Confidence limits on phylogenies: An approach using the bootstrap. *Evolution,* 39, 783-791.

Fields, B.N., Knipe, D.M., Howley, P.M., Griffin, D.E. (2001): Fields virology. 4th ed. Philadelphia: Lippincott Williams & Wilkins.

Goebel, S.J., Miller, T.B., Bennett, C.J., Bernard, K.A., Masters, P.S. (2007): A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *Journal of virology,*81: 1274–1287.

Hall, T.A., (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41: 95-98.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al (2020): Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet,* 395: 497–506.

Kaur, N., Singh, R., Dar, Z., Bijarnia, R. K., Dhingra, N., Kaur, T. (2020): Genetic comparison among various coronavirus strains for the identification of potential vaccine targets of SARS-CoV2. Infection, genetics and evolution. journal of molecular epidemiology and evolutionary genetics in infectious diseases 104490. Advance online publication. https://doi.org/10.1016/j.meegid.2020.104490.

Lai, M.M., and Cavanagh, D. (1997): The molecular biology of coronaviruses. *Advances in virus research,*48: 1–100.

Lai, M.M., Perlman, S., Anderson, L. (2007): Coronaviridae. In Fields Virology. Knipe, D.M., Howley, P.M., Eds., Lippincott Williams & Wilkins: Philadelphia, PA, USA v1: 1305–1318.

Lai, M.M., and Stohlman S.A. (1981): Comparative analysis of RNA genomes of mouse hepatitis viruses. *Journal of Virology,* 38: pp. 661-670.

Li, T., Zhang, Y., Fu, L., Yu, C., Li, X., Li, Y., Zhang, X. et al. (2005): siRNA targeting the leader sequence of SARS-CoV inhibits virus replication. *Gene therapy,*12(9):751-61. doi: 10.1038/sj.gt.3302479.

Li, L., Kang, H., Liu, P., Makkinje, N., Willi amson, S. T., Leibowitz, J. L., Giedroc, D. P. (2008): Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication. *Journal of molecular biology*,377(3):790-803.

Madhugiri, R., Karl, N., Petersen, D., Lamkiewicz, K., Fricke, M. et al. (2018): Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology,* 517:44-55.

Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield Y.S. et al. (2003): The genome sequence of the SARS-associated coronavirus. *Science*, 300:1399–1404.

Nei, M., and Kumar S. (2000): Molecular Evolution and Phylogenetics. Oxford

University Press,Oxford, England/ New York, USA, 333 pp.

Paraskevis, D., Kostaki, E.G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., Tsiodras, S. (2020): Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, genetics and evolution,*79, 104212.

Perlman, S., and Netland J. (2009): Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology,* 7:439-450. doi: 10.1038/nrmicro2147.

Qin, E., Zhu, Q., Yu, M. Li Y, Wu Q., Lin W., Chen W., Tang L. et al. (2003a): A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01). *Chinese science bulletin,*48: 941–948. https://doi.org/10.1007/ BF03184203.

Qin E., He X., Tian W., Liu Y., Li W., Wen J., Wang J., Fan B., Wu Q. et al. (2003b): A genome sequence of novel SARS-CoV isolates: the genotype, GD-Ins29, leads to a hypothesis of viral transmission in South China. *Genomics Proteomics Bioinformatics,*1: 101-107. doi: 10.1016/s1672-0229(03)01014-3.

Rangan, R., Zheludev I.N., Das R. (2020): RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses. bioRxiv. //doi.org/10.1101/2020.03.27.012906

Rota, P.A, Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., PeñarandaS. et al. (2003): Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science,* 300: 1394–1399.

Sawicki S.G., and Sawicki D.L. (1998): A new model for coronavirus transcription. *Advances in experimental medicine and biology*, 440: 215–219.

Sola, I., Almazán, F., Zúñiga, S., Enjuanes, L. (2015): Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual review of virology,* 2: pp. 265-288

Tamura, K., Nei, M., Kumar, S. (2004): Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences*, (USA), 101 (30): 11030-11035. DOI: 10.1073/pnas.0404206101.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar S. (2013): MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution,* 30: 2725-2729.

Weiss, S.R., and Navas-Martin, S. (2005): Coronavirus pathogenesis and the emerging pathog syndrome coronavirus. *Microbiology and molecular biology reviews,* 69: 635–664.

Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y. (2009): Coronavirus Diversity, Phylogeny and Interspecies Jumping. *Experimental biology and medicine,* (Maywood) 234:1117-1127. doi: 10.3181/0903-MR-94.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J. (2020): Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host and microbe,*27: 325–328.

Yang, D., and Leibowitz, J.L. (2015): The structure and functions of coron virus genomic 3’ and 5’ ends. *Virus research*, 206:120e33.

Zeng, F.Y., Chan, C.W., Chan, M.N., Chen, J.D., Chow, K.Y., Hon, C.C., et al. (2003): The complete genome sequence of severe acute respiratory syndrome coronavirus strain HKU-39849 (HK-39). *Experimental biology and medicine,* (Maywood) 228(7):866-873.

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

X.F. (2006): Promoter activity of SARS coronavirus 5' UTR sequence in eukaryotic cells. *Sichuan Da XueXue Bao Yi Xue Ban,*37(1): 5-9. PMID: 16468630.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L. (2020): Apneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature,* 579: 270–273.

Zust, R., Miller, T.B., Goebel, S.J., Thiel, V., Masters, P.S. (2008): Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *Journal of virology,* 82 (3): 1214–1228.

ARABIC SUMMARY

التتابع النيكليوتيدى و التنوع فى المناطق الجينومية 5'UTRs للفيروسات التاجية (betacronoviruses)

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

فى هذه الدراسة تم تحليل المناطق الجينومية 5'UTRs للفيروسات التاجية فيما يتعلق باختلافات التسلسل النيكليوتيدى ومكوناته و كذلك أنماط العلاقة التطورية الوراثية.
أوضحت الدراسة ان تلك المناطق (5'UTRs)محفوظة بدرجة كبيرة داخل المادة الوراثية الفيروسية و الواضحه بين الفيروسات وثيقة الصلة، كما اشارت لاحتواء تلك المناطق لمواقع متغيرة من التتابع النيكليوتيدى.

أظهرت الدراسة وجود ثلاثة سترات محفوظة ذات تتابع نيوكليوتيدى طويل فى تلك المنطقة كالاتى: nt61 to nt107، nt211- nt236 و ايضا nt262 to nt277، من بينهم السترة (61nt إلى nt107) ذات التتابع النيكليوتيدى الطويل بها ما يقرب من نصف تسلسل leader sequence (61nt إلى nt93) مما يؤكد وجود هذا التسلسل في الجينوم الفيروسي .كما اظهرت ايضا' معظم ان المحتوى النيكليوتيدى للمنطقة (5'UTR) من الأدنينAT) ثيامين/) مقارنة بالجوانين/سيتوسين (GC)، والذى ربما يكون اساسيا للنشاط الفيروسى و نسخه المتماثل.

بينت الدراسة الآختلاف فى حجم مناطق 5'UTRs للفيروسات التاجية المدروسة من حيث التتابع النيكليوتيدى الذى تراوح من 245 زوج قاعدة لفيروس سارس BJ01 الى 238 زوج قاعدة لفيروسhCoV-19_Chile ' و هذا الآخير احتوى على 18 زوج قاعدة اضافية مقارن بالفيروسات الاخرى.

أكد التحليل الوراثي باستخدام طريقة الآحتمال القصوى لتلك المناطق (5'UTRs) المسافة التطورية القريبة والآصل المشترك للفيروسات التاجية (betacronoviruses) حيث كان اقرب الفيروسات وراثيا للفيروس التاجى SARS-CoV-2 Wuhan-Hu-1 هو فيروس الخفاش التاجي RaTG13' و ايضا بعدا تطوريا لفيروسى السارس التاجية (BJ01 and GD01) عن الفيروس التاجى SARS-CoV-2 مما اكد ظهورسلالتهم المبكر.

كشف التجميع الوراثى التطوري فى مناطق (5'UTRs) للفيروسات التاجية (betacronoviruses) عن التنوع الوراثى و حدوث الطفرات الجينية و الاستنساخ الفيروسى المتجدد' الذى ربما يشير إلى زيادة خطر انتقال العدوى بين الأنواع و و المجتمعات.