Central Asian Journal of Medicine

THE ROLE OF ORF1AB AND N GENES IN UNDERSTANDING THE SARS-COV-2 RNA SEQUENCE (Literature Review)

Sobirjon T. Jumamurodov [1], Zukhra A. Nuruzova [2]

1 PhD candidate of the Department of Microbiology, Virology and Immunology, Tashkent Medical Academy, Tashkent, Uzbekistan
E-mail: sobirjonjumomurodov@gmail.com

2 Professor of the Department of Microbiology, Virology and Immunology, Tashkent Medical Academy, Tashkent, Uzbekistan
E-mail: voyna_microbam@mail.ru

ABSTRACT

Molecular diagnostic methods are widely used in the detection of the SARS-CoV-2 virus, in particular, real-time polymerase chain reaction (RT-PCR) technology is considered the gold standard for these purposes. Various genetic markers in the viral genome are used to identify them, of which the ORF1ab and N genes are the most important. These genes contain sequences that are specific to the virus, and by targeting them, SARS-CoV-2 RNA can be identified with high sensitivity and accuracy. The article is devoted to the analysis of the scientific literature and role of genes ORF1ab and N in molecular diagnostics.

Key words: SARS-CoV-2, ORF1ab gene, N gene, RNA testing, virus detection, molecular diagnostics, polymerase chain reaction (PCR), genetic assay, virus genomics, COVID-19.

INTRODUCTION

During the first month of the SARS-CoV-2 epidemic, the rapid development of PCR-based diagnostic tests became a global priority, therefore timely diagnosis, isolation and contact tracing could minimize the impending surge of the pandemic. The development of these tests for widespread, long-term detection has been complicated by limited data on the genome sequence of a new virus and how it can mutate as it spreads globally and adapts to humans [1].

The ability to detect SARS-CoV-2 in a widespread epidemic is critical to the success of carrier inspection and quarantine efforts. Reverse transcription-polymerase chain reaction (RT-qPCR) and real-time sequencing-based methods

are used to identify and characterize viruses. However, RNA viruses are known for their high genetic diversity, which makes it difficult to develop effective nucleic acid assays. The first genomic sequencing of SARS-CoV-2 revealed new mutations that may already affect the performance of existing screening tests, leading to mis-negative diagnoses or ineffective treatments [2].

At the end of 2019, a new coronavirus appeared in humans - SARS-CoV-2. Phylogenetic evidence points to the zoonotic origin in Wuhan, the capital of central China's Hubei province, from where the new virus quickly spread around the world and became a pandemic. SARS-CoV-2 belongs to the genus β-coronaviruses of the family Coronaviridae and has been linked to other viruses that cause infections in humans, such as SARS-CoV and MERS-CoV. The new SARS-CoV-2 is 80% similar to SARS-CoV (the causative agent of the SARS outbreak in Asia in 2002-2003) and almost 96% similar to the isolation of the bat coronavirus RaTG13, indicating that these animals are a natural reservoir. The SARS-CoV-2 genome consists of a single positive-stranded RNA consisting of approximately 30,000 nucleotides. As the epidemic intensified, multiple genomic sequences became available in public databases by researchers around the world. The high resilience and infection of RNA viruses is due in part to their high mutation rate, which spread over a two-and-a-half month period from unknown cases of pneumonia in Wuhan, China, to the World Health Organization. WHO) - declared a state of emergency of a global pandemic, covering 114 countries, during which the accelerated development of laboratory diagnostics of the new virus became a global priority; Rapid diagnosis, isolation and contact tracing have become central parts of controlling the impending pandemic epidemic. Several government, hospitals, and academic laboratories have developed PCR analyses compiled by the World Health Organization (WHO) for global distribution. Seven of these laboratory-developed tests (LDTs) were published on the WHO website on January 24, 2020, a month after the first reported cases of the disease. Additional LDT was developed in mid-January 2020 in the public health laboratory of the Centre for Disease Control of British Columbia [3] [4] [5].

PCR-based diagnostic tests rely on the careful design of synthetic oligonucleotide primers and probes. Nucleotide mismatch between primers, probes, and target genetic material leads to thermodynamic instability that can disrupt PCR chemistry, disrupt detection, and lead to false-negative results [6]. This is especially problematic in detecting RNA viruses, such as coronaviruses, whose genomes easily mutate compared to DNA-based organisms. Consequently, oligonucleotide strategies for the development of viral pathogens focus on genomic

loci where low mutation rates are critical to maintaining biological function and pathogen viability [7] [8].

Newly emerging pathogens present unique challenges for the development of oligonucleotides because their genomes are poorly characterized. Without extensive genome sequencing, it would be difficult to isolate stable genomic loci that can be distinguished by oligonucleotide design. For nascent zoonotic viruses, it is difficult to predict the genomic changes that these problems would cause with adaptation to a new host. This makes the pandemic caused by the novel zoonotic RNA virus a challenging scenario for the development of PCR-based diagnostics: oligonucleotide design choices pose significant risks, but should be made with incomplete and insufficient data [9].

The SARS-CoV-2 coronavirus contains a linear, single-stranded positive RNA genome. The SARS-CoV-2 coronavirus genome consists of the ORF1ab instruction sequence, which codes for proteins for RNA replication and non-structural proteins (NSPs) and structural proteins. The genomic guidance sequence around BP 265 is a feature of coronavirus replication and plays an important role in coronavirus gene expression during its continuous subgenomic replication [10]. ORF1ab encodes replicase polyproteins required for viral RNA replication and transcription. Ribosomal scaffolding shifting is required to represent the c-proximal portion of ORF1ab (−1). The first non-structural protein (nsp) encoded by ORF1ab is papain-like proteinase (PL proteinase, nps3). Nsp3 is the most important and largest component of the replication and transcription complex. PROTEINASE PL 1-3 breaks down non-structural proteins and blocks the host's innate immune response, stimulating cytokine expression. Nsp1ab-encoded in ORF4ab is responsible for the formation of double membrane bubbles (DMVs) [11]. Other non-structural proteins are 3CLPro (3-chymotrypsin-like proteinase, 3CLpro) and nsp6 proteinases. The 3CLPro protease is required for RNA replication. Proteinase 3CLPro is responsible for the processing of the nsp4 terminal via nsp16 in all coronaviruses. Thus, the preserved structure and catalytic sites of 3CLpro could serve as attractive targets for antiviral drugs. nsp3, nsp4, and nsp6 together can induce DMV [12].

SARS coronavirus RNA replication is rare, including two RNA-linked RNA polymerases (RNA POL). The first RNA polymerase is a primer-dependent structural protein 12 (nsp12), while the second RNA polymerase is nsp8. Unlike nsp12, nsp8 has the ability to initiate de novo replication without primers. Nsp7 and nsp8 play important roles in SARS-CoV-2 replication and transcription. The SARS-coronavirus nsp7-nsp8 complex is a multimeric RNA polymerase for de novo initiation and primary expansion. Nsp8 also interacts with an additional

protein, ORF6. The SARS coronavirus replicase protein NSP9 binds RNA and interacts with nsp8 for its functions [13].

In addition, the SARS-CoV-2 genome encodes four component proteins. Structural proteins exhibit a significantly higher immunogenicity to T-cell reactions than nonstructural proteins. Structural proteins are involved in various viral processes, including the formation of viral particles. Structural proteins include the spike (C), envelope (E), membrane protein (M), and nucleoprotein (N), which are common to all coronaviruses. Spike C protein is a glycoprotein having two domains, S1 and S2. The S1 spike protein attaches the virus to the cell membrane and interacts with the host receptor ACE2 to initiate infection [14]. After introducing the virus into the host cell endosomes, glycoprotein C is stimulated by conformational changes. The C protein is then cleaved by catepsin CTSL and it opens the thermonuclear peptide S2 and thereby activates membrane synthesis in endosomes. The S2 domain of the spike protein mediates the synthesis of virion and cell membranes, in particular, the spike glycoprotein of the SARS-CoV-2 coronavirus contains a furin-like fragmentation site. The place of recognition of furin is important for its recognition by pyrolysis and thus to facilitate the transmission of zoonotic virus infection [15]. The envelope protein (e) interacts with membrane protein M in the bud compartment of the host cell. Protein M has dominant cell immunogenicity. The nucleoprotein (ORF9a) packs the positive strand of the viral RNA genome into a spiral ribonucleocapsid (RNP) during virion deposition through its interaction with the viral genome and membrane protein M. The nucleoprotein plays an important role in enhancing the effectiveness of subgenomic viral RNA. Transcription is also known as viral replication [16] [17].

Increasing epidemiological and clinical evidence suggests that SARS-CoV-2 has higher transmissibility and lower pathogenicity than SARS-CoV. However, the mechanism of high prevalence of SARS-CoV-2 remains unclear. DNA sequence comparisons using single nucleotide polymorphisms (SNPs) are often used for evolutionary studies and may be particularly useful in recognizing mutated coronavirus genomes, where high mutations due to error-prone RNA-dependent RNA polymerase can occur during genome replication [18].

The SARS-CoV-2 pandemic has caused great health and economic stress in the world. Therefore, understanding the nature of this virus and developing methods to control the spread of the virus during a pandemic is critical for disease control. The results reveal several molecular aspects of SARS-CoV-2 related to this pandemic [19] [20].

In conclusion, this article provides a scientific basis on a scientific basis the role of molecular diagnostics, in particular RT-PCR technology, in detecting SARS-CoV-2. The study analyzed the role of genes ORF1ab and N in the viral genome for diagnosis, revealing the important functions of these genes for viral replication and the formation of structural proteins. Targeting these genetic markers plays an important role in effectively managing the pandemic, improving diagnostic accuracy, and reducing the likelihood of false positives. These studies make a significant scientific contribution to the improvement of diagnostic and treatment strategies for pandemic management.

REFERENCES

1. Andersen, K. G., et al. "The proximal origin of SARS-CoV-2." *Nature Medicine*, 26(4), 2020.
2. Chan, J. F., et al. "Improved molecular diagnostics for SARS-CoV-2 detection." *Journal of Clinical Microbiology*, 2020.
3. Corman, V. M., et al. "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR." *Euro Surveill*, 25(3), 2020.
4. Drosten, C., et al. "Identification of a novel coronavirus in patients with severe acute respiratory syndrome." *New England Journal of Medicine*, 348(20), 2003.
5. Gorbalenya, A. E., et al. "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2." *Nature Microbiology*, 5, 2020.
6. Hoffmann, M., et al. "SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor." *Cell*, 181(2), 2020.
7. Hu, B., et al. "Characteristics of SARS-CoV-2 and COVID-19." *Nature Reviews Microbiology*, 19(3), 2021.
8. Kim, D., et al. "The architecture of SARS-CoV-2 transcriptome." *Cell*, 181(4), 2020.
9. Kuo, L., & Masters, P. S. "The small envelope protein E is critical for coronavirus assembly." *Journal of Virology*, 76(8), 2002.
10. Li, F. "Structure, function, and evolution of coronavirus spike proteins." *Annual Review of Virology*, 3, 2016.
11. Liu, Y., et al. "The N protein of coronavirus in diagnostics and therapeutics." *Virus Research*, 2020.

12. Lu, R., et al. "Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding." *The Lancet*, 395(10224), 2020.

13. Perlman, S., & Netland, J. "Coronaviruses post-SARS: update on replication and pathogenesis." *Nature Reviews Microbiology*, 7(6), 2009.

14. van Dorp, L., et al. "Emergence of genomic diversity and recurrent mutations in SARS-CoV-2." *Infection, Genetics and Evolution*, 83, 2020.

15. Walls, A. C., et al. "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein." *Cell*, 181(2), 2020.

16. World Health Organization. "Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: interim guidance." WHO, 2020.

17. Wrapp, D., et al. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." *Science*, 367(6483), 2020.

18. Xu, J., et al. "Epidemiological data on SARS-CoV-2 and molecular analysis of spike protein evolution." *Emerging Infectious Diseases*, 26(7), 2020.

19. Zhou, P., et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin." *Nature*, 579, 2020.

20. Zhu, N., et al. "A novel coronavirus from patients with pneumonia in China, 2019." *New England Journal of Medicine*, 382(8), 2020.