Short communications

Less frequent sequence mismatches in variants of concern (VOCs) of SARS-CoV-2 in the real-time RT-PCR assays developed by the National Institute of Infectious Diseases, Japan

Kazuya Shirato¹*, Shutoku Matsuyama¹,², and Makoto Takeda¹

¹Department of Virology III, and ²Influenza and Respiratory Virus Research Center, National Institute of Infectious Disease, 4-7-1 Gakuen, Musashimurayama, Tokyo 208-0011, Japan

Running title: Mismatches in NIID PCR assays for SARS-CoV-2

Keywords: Coronavirus Diseases 2019 (COVID-19), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), variant of concern (VOC), real-time RT-PCR

*Corresponding author:
Kazuya Shirato, DVM, PhD.
Chief, Laboratory of Respiratory Virus Vaccines
Department of Virology III
National Institute of Infectious Diseases, Murayama Branch
4-7-1 Gakuen, Musashimurayama
Tokyo, 208-0011, Japan
E-mail: shirato@nih.go.jp
Tel: +81-42-561-0771
Fax: +81-42-567-5631
著者一覧
白戸憲也 〒208-0011 東京都武蔵村山市学園 4-7-1
国立感染症研究所ウイルス第3部
松山州徳 〒208-0011 東京都武蔵村山市学園 4-7-1
国立感染症研究所ウイルス第3部
国立感染症研究所インフルエンザ・呼吸器系ウイルス研究センター
竹田誠 〒208-0011 東京都武蔵村山市学園 4-7-1
国立感染症研究所ウイルス第3部

責任著者連絡先
白戸憲也
〒208-0011 東京都武蔵村山市学園 4-7-1
国立感染症研究所ウイルス第3部
Tel: 042-561-0717
Fax: 042-567-5631
E-mail: shirato@nih.go.jp
Summary

Various variants of severe acute respiratory syndrome (SARS) coronavirus 2 (SARS-CoV-2) began emerging worldwide from the end of 2020 to the beginning of 2021. The variants GRY/VOC202012/01 (B1.1.7), GH/N501Y.V2 (B1.351), and GR/N501Y.V3 (P1) are characterized by N to Y amino acid substitution at position 501 in the S protein. The variant containing L to R substitution at position 452 in the S protein G/L452R.V3 (B1.617) was endemic to India. The heightened concern regarding these variants is related to their increased viral infectivity. Information about nucleotide mismatch(es) on the primer/probe sequence is important for maintaining good performance of real-time PCR assays. In this study, real-time RT-PCR assays developed by the National Institute of Infectious Diseases, Japan (NIID-N2 and NIID-S2 assays), were reviewed to analyze nucleotide mismatches of variants in primer/probe sequences. The frequency of mismatched sequences in three variants (GRY/VOC202012/01, GH/N501Y.V2, and GR/N501Y.V3) was lower than that in all SARS-CoV-2 sequences. The mismatch, that G to C substitution at nucleotide 8 in reverse primer of S2 set, elevated to about 16.3% in G/L452R.V3, however the substitution did not affect the analytical sensitivity of assay. Therefore, the study indicates that the NIID-N2 and NIID-S2 sets detect VOCs of SARS-CoV-2 with reliable efficiency.
The coronavirus disease 2019 (COVID-19) outbreak caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) originated in Wuhan, China, in December 2019, and has rapidly spread worldwide (1). As of May 25, 2021, there have been 167,011,807 confirmed cases and 3,472,068 deaths reported globally (2).

As soon as the viral sequence was published (GenBank NM908947), the National Institute of Infectious Disease, Japan (NIID) developed a real-time reverse transcription-polymerase chain reaction (RT-PCR) assay targeting the viral nucleocapsid (N) protein region for national surveillance of SARS-CoV-2; this RT-PCR assay was named the NIID-N2 assay (3). The NIID subsequently developed a real-time RT-PCR targeting the spike (S) protein region, which was designated NIID-S2 assay, in preparation for the possible emergence of variants with reduced sensitivity by NIID-N2 assay (4). Thus, both NIID-N2 and NIID-S2 assays are available for national surveillance of SARS-CoV-2 in Japan. The primers/probe sequences are shown in Table 1 (showed as “original”).

Information about nucleotide mismatch(es) on the primer/probe sequence is essential for maintaining good performance of the real-time PCR assays. Therefore, we recently reported the one-year tendency of outstanding mismatches in primer/probe target region used in both NIID-N2 and NIID-S2 assays (4). The report described a C to T
substitution at nucleotide (nt)20 (N2F_C20T) in the forward primer of NIID-N2 assay that affected the analytical sensitivity of the assay and a G to T substitution at nt6 (N2P_G6T) of the probe that possibly affected the slope of the amplification curve, without affecting the analytical sensitivity, owing to the combination of real-time PCR instrument and reagents used (4).

Variants of SARS-CoV-2 began emerging globally from the end of 2020 to the beginning of 2021. The variant B.1.1.7 lineage (GRY/variant of concern [VOC] 202012/02) emerged in England in November 2020 and rapidly spread worldwide (5). The variant B.1.351 lineage (GH/N501Y.V2) independently emerged in South Africa (6). The variant P.1 lineage (GR/N501Y.V3) was also emerged and detected in Brazilian travelers at a quarantine depot in Haneda Airport, Japan (7). These three variants were characterized by N to Y amino acid substitution at position 501 in the S protein, which may promote increased transmission (8). The 501Y.V2 and 501Y.V3 variants carry the E to K substitution at position 484 in the S protein, resulting in decreased neutralization activities of recovered patients’ or vaccinees’ sera (9, 10). Furthermore, the variant B1.617 lineage containing the L to R substitution at position 452 in the S protein (G/L452R.V3) was first detected in India (11), and this mutation might result in increased viral infectivity and immune evasion in the host (12, 13). Considering the
importance to detection, nt mismatches to these variants in the primer/probe sequences of NIID-N2 and NIID-S2 assays was evaluated, whether these assays can detect them without reducing the efficiency.

As previously described (4), quality sequence data submitted by May 9, 2021, were obtained from the GISAID database (https://www.gisaid.org/) by searching “complete”, “high coverage”, “low coverage excl”, and “human”. The obtained sequences were aligned with the Wuhan-Hu-1 isolate (GenBank: MN908947) using a multiple alignment program for amino acid or nucleotide sequences (MAFFT version 7, https://mafft.cbrc.jp/alignment/server/add_fragments.html?frommanual) (14). Mismatches were analyzed using SEQUENCHER software (Gene Codes, Ann Arbor, MI). The number of mismatches for all available sequences are listed in Table 2, and their time-courses are presented in Fig. 1. By May 9, 2021, approximately 1,060,000 sequences were available, of which 46% were variants (Table 2). As of March 7, 2021, approximately 500,000 sequences were available, of which 20% were VOC 202012/02, suggesting explosion of the VOC sequence deposition in just 2 months. As cautionary point, there was the bias for sequence origin in GISAID database identical to previous report (4), because about 90% of sequences were registered from Europe and North America.

6
The list of outstanding mismatches, which are detected in more than 400 sequences, including reported in a previous study (4) are shown in Table 1. In addition to previous report, C to T substitution at nt13 (N2F_C13T) and G to A substitution at nt16 in the NIID-N2 forward primer (N2F_G16A), C to T substitution at nt17 in the NIID-N2 reverse primer (N2R_C17T), T to C substitution at nt12 (SF1_T12C), T to C substitution at nt18 (SF1_T18C) in the NIID-S2 forward primer, G to A substitution at nt6 (SR3_G6A), G to A, and G to C substitution at nt8 (SR3_G8A and SR3_G8C) in NIID-S2 in reverse primer were identified as new mismatches. In all data, 2.49% mismatches were observed in the NIID-N2 set and 1.83% mismatches were observed in the NIID-S2 set (Table 1). By the end of 2020, the N2P_G6T in the United Kingdom had the most outstanding mismatches and occupied more than half of the total mismatches in the NIID-N2 set (4); however, in 2021, the prevalence of N2P_G6T decreased, whereas those of other mismatches increased (Fig 1). The entire mismatch rate of in the S2 set increased as compared with that reported previously (4). It was due to increased prevalence of SF1_T18C, SP2_G2T, and SP2_T22C. SF1_T18C was almost entirely specific to Canada, while SP2_T22C was reported mostly in Denmark. Although SP2_G2T spread widely, the number of registrations from the USA was notable. The effect of newly identified mismatches for amplification efficiencies were
evaluated using a synthesized mutated RNA template, as previously described (4). N2F_C13T, N2F_G16A and SF1_T18C decreased the analytical sensitivities by 10- to 20-fold, although others did not affect the detection efficiency, showing fewer than ten copies of analytical sensitivities (Table 1). N2F_20T mismatches decreased the analytical sensitivity identical to previous report (4). The top country of registration of N2F_20T was Hong Kong (653, 24.2%) and it occupied 37.1% of total submission from Hong Kong (1758 sequences). Although the mismatch rate of N2F_20T in VOC variants is lower, the result suggests that it requires special care in the detection of specimen derived from Hong Kong using NIID-N2 set. On the other hand, the frequency of mismatched sequences in the three N501Y variants was lower as compared to all data. Exceptionally, the existence rate of N2Rver3_G11T was higher in B1.1.7 variants (0.68%) than that in all available sequences (0.52%) (Table 2), but it did not affect the analytical sensitivity of NIID-N2 assay (Table 1). In contrast, the rate of SR3_G8C mismatch was characteristically shown in G/L452R.V3 VOC and it occupied most of mismatches seen in the G/L452R.V3 (Table 2). More than half of SR3_G8C mismatch was registered from India and it accounted for 22.3% of total registered G/L452R.V3 VOC sequences by India (287 of 1288). Although SR3_G8C accounted for 16.3% of mismatches seen in G/L452R.V3 VOC, it did not affect the analytical
sensitivity (Table1), indicating NIID-S2 set was available for the detection of G/L452R.V3 VOC.

Most of mismatches are single substitution in the sequences targeted by NIID-N2 and NIID-S2 primer/probe sets and the sequence that had more than two mismatches were less frequent in all mismatched sequences (0.37%, NIID-N2; 0.30%, NIID-S2). The results here showed the amplifications were not abolished completely by single mismatches in both NIID-N2 and NIID-S2 sets. These data indicate that the NIID-N2 and NIID-S2 sets can detect VOC variants of SARS-CoV-2 without concerns of low detection efficiency.

The available sequence numbers for GRY/ VOC202012/01 have reached more than 460,000 in 5 months (Fig. 1). The N501Y VOCs is going on replacing the currently predominant SARS-CoV-2 population owing to their fast transmission characteristics. It is, therefore, critical to monitor the prevalence of nt mismatches in RT-PCR assays for N501Y VOCs.

**Acknowledgments**

We thank all members of GISAID for deposits of valuable SARS-CoV-2 sequences.
Conflict of interest

None to declare.

Funding

This work was supported by Grants-in-Aid (nos. 19fk0108110j0001 and 20fk0108117j0101) from the Japan Agency for Medical Research and Development, by Grant-in-Aid from the Japanese Ministry of Health Labour Sciences Research (20HA2007), and by Grant-in-Aid from the Japan Society for the Promotion of Science (C:20K06441).

References

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265-9.

2. WHO. WHO Coronavirus Disease (COVID-19) Dashboard. Available at: <https://covid19.who.int/>. Accessed Oct. 21 2020.

3. Shirato K, Nao N, Katano H, et al. Development of Genetic Diagnostic Methods for Detection for Novel Coronavirus 2019(nCoV-2019) in Japan. Jpn J Infect Dis. 2020;73:304-7.
4. Shirato K, Tomita Y, Katoh H, et al. Performance evaluation of real-time RT-PCR assays for detection of severe acute respiratory syndrome coronavirus-2 developed by the National Institute of Infectious Diseases, Japan. Jpn J Infect Dis. 2021. DOI: 10.7883/yoken.JJID.2020.1079.

5. Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. 2021.

6. Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. MedRxiv. 2020.

7. National Institute of Infectious Diseases J. Brief report: New Variant Strain of SARS-CoV-2 Identified in Travelers from Brazil. Available at: <https://www.niid.go.jp/niid/images/epi/corona/covid19-33-en-210112.pdf>. Accessed Mar 9 2021.

8. Leung K, Shum MH, Leung GM, et al. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill. 2021;26.

9. Jangra S, Ye C, Rathnasinghe R, et al. The E484K mutation in the SARS-CoV-2 spike protein reduces but does not abolish neutralizing activity of human convalescent
and post-vaccination sera. medRxiv. 2021.

10. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature. 2021.

11. Latif AA, Mullen JL, Alkuzweny M, et al. India Mutation Report. Available at: <https://outbreak.info/location-reports?loc=IND>. Accessed 26 May 2021.

12. Tchesnokova V, Kulakesara H, Larson L, et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. BioRxiv. 2021.

13. Yadav PD, Sapkal GN, Abraham P, et al. Neutralization of variant under investigation B.1.617 with sera of BBV152 vaccinees. Clin Infect Dis. 2021.

14. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160-6.
Figure legends

Figure 1. Time-course of the number of registered sequences. Upper graphs indicate NIID-N2 assay; lower graphs indicate NIID-S2 assay. Left panels show available sequence numbers; middle panels show number of mismatches; right panels show occurrence rate of mismatches against total number of available sequences.
Table 1. List of mismatches in primer/probe sequences of NIID-N2 and NIID-S2 sets and their effects to analytical sensitivities.

| NIID-N2 set | Position of mismatches in primer/probe sequence (5' to 3') | Analytical sensitivity (copy numbers)* |  |  |  |
|-------------|-------------------------------------------------------------|---------------------------------------|---|---|---|
|             |                                                             | Original set                            | N2F_C13T was used as F primer | N2F_G16A was used as F primer | N2F_C20T was used as F primer |
| NIID_2019-nCOV_N_F2 (original) | AAATTATTGGGGACCAGGAAC | 1.4 | N/A | N/A | N/A |
| N2F_C13T    | AAATTATTGGGA**CAGGAAC | 140.6 | 7.9 | N/A | N/A |
| N2F_G16A    | AAATTATTGGGACCA**GAAC | 25.0 | N/A | 2.5 | N/A |
| N2F_G16T    | AAATTATTGGGACCA**CAAC | 4.4 | N/A | N/A | N/A |
| N2F_C20T    | AAATTATTGGGACCA**GGAAC | 79.1 | N/A | N/A | 7.9 |
| NIID_2019-nCOV_N_R2ver3 (original) | TGGCACCTGTGTAGGTCAAC | 1.4 | N/A | N/A | N/A |
| N2Rver3_G9A | TGGACCTATGTAGGTCAAC | 7.9 | N/A | N/A | N/A |
| N2Rver3_G11T| TGGACCTTAGGTCAAC | 4.4 | N/A | N/A | N/A |
| N2Rver3_C17T| TGGACCTGTAGGTCAAC | 1.4 | N/A | N/A | N/A |
| NIID_2019-nCOV_N_P2 (original) | ATGTGCATATGGCATGGA | 1.4 | N/A | N/A | N/A |
| N2P_G6T     | ATGC**TGCAATTGGCATGGA | 2.5 | N/A | N/A | N/A |

| NIID-S2 set | Position of mismatches in primer/probe sequence (5' to 3') | Analytical sensitivity (copy numbers)* |  |  |
|-------------|-------------------------------------------------------------|---------------------------------------|---|---|
|             |                                                             | Original set                            | SF1T18C was used as F primer |
| SARS-CoV2_NIID_S_F1 (original) | CAGTCAGCACCCTCATGTTGA | 2.5 | N/A |
| SF1_T12C    | CAGTCAGCACCCTCATGTTGA | 0.8 | N/A |
| SF1_T18C    | CAGTCAGCACCCTCATGTTGA | 44.4 | 2.5 |
| SARS-CoV2_NIID_S_R3 (original) | AACCAGTGTTGTCATTTGA | 2.5 | N/A |
| SR3_G6A     | AACCAGTGTTGTCATTTGA | 2.5 | N/A |
| SR3_G8A     | AACCAGTGTTGTCATTTGA | 1.4 | N/A |
| SR3_G8C     | AACCAGTGTTGTCATTTGA | 4.4 | N/A |
| SR3_T9C     | AACCAGTGTTGTCATTTGA | 4.4 | N/A |
| SARS-CoV2_NIID_S_P2 (original) | TGCTCCTGCTATTTGATTTGA | 2.5 | N/A |
| SP2_G2T     | TGCTCCTGCTATTTGATTTGA | 1.4 | N/A |
| SP2_G20T    | TGCTCCTGCTATTTGATTTGA | 4.4 | N/A |
| SP2_T22C    | TGCTCCTGCTATTTGATTTGA | 4.4 | N/A |

* The analytical sensitivities were evaluated by the original N2 and S2 sets using serially diluted template that each mismatch was introduced. To confirm the effect of mismatches, the detection was performed replacing the primer to that contained the mismatch. The analytical sensitivities were calculated by Reed-Muench methods and were expressed as real copy numbers.
Table 2. Analysis of mismatches in the sequences of NIID assay primers/probes.

| NIID-N2 set | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches |
|-------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|
|             | All*    | Rate (%) | GRY/ VOC202012/01 (B1.1.7) | Rate (%) | GH/N501Y.V2 (B1.351) | Rate (%) | GR/N501Y.V3 (P.1) | Rate (%) | G/L452R.V3 (B1.617) | Rate (%) | Top reported country (%) |
| N2F_13T     | 1066262 | 2.49     | 459145 | 1.52 | 11855 | 0.84 | 11159 | 0.5 | 2777 | 0.94 | Germany (51.2) |
| N2F_16T     | 601     | 0.06     | 467   | 0.1 | 0    | 0 | 1 | 0.01 | 0 | 0 | USA (65.3) |
| N2F_16A     | 615     | 0.06     | 113   | 0.02 | 25 | 0.21 | 10 | 0.09 | 1 | 0.04 | USA (30.3) |
| N2F_20T     | 672     | 0.06     | 63    | 0.01 | 0    | 0 | 0 | 0 | 0 | 0 | USA (65.3) |
| N2Rver3_G9A | 2038    | 0.19     | 386   | 0.08 | 3   | 0.03 | 1 | 0.01 | 0 | 0 | Hong Kong (24.2) |
| N2Rver3_G11T| 1924    | 0.18     | 663   | 0.14 | 31  | 0.26 | 3  | 0.03 | 4 | 0.14 | England (32.8) |
| N2Rver3_C17T| 5552    | 0.52     | 3117  | 0.68 | 8   | 0.07 | 6 | 0.05 | 2 | 0.07 | Japan (24.2) |
| N2P_G6T     | 3859    | 0.36     | 3     | 0 | 1 | 0.01 | 0 | 0 | 3 | 0.11 | USA (98.7) |

| NIID-S2     | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches | Available sequences | Mismatches |
|-------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|---------------------|------------|
|             | All*    | Rate (%) | GRY/ VOC202012/01 (B1.1.7) | Rate (%) | GH/N501Y.V2 (B1.351) | Rate (%) | GR/N501Y.V3 (P.1) | Rate (%) | G/L452R.V3 (B1.617) | Rate (%) | Top reported country (%) |
| SF1_T12C    | 19504   | 1.83     | 2533  | 0.55 | 78  | 0.66 | 17 | 0.15 | 474 | 16.99 | Germany (35.7) |
| SF1_T18C    | 1162    | 0.11     | 389   | 0.08 | 0   | 0 | 0 | 0 | 0 | 0 | Canada (93.0) |
| SP2_G2T     | 1231    | 0.12     | 43    | 0.01 | 0   | 0 | 0 | 0 | 0 | 0 | USA (18.3) |
| SR3_G8A     | 2818    | 0.26     | 308   | 0.07 | 13  | 0.11 | 8 | 0.07 | 1 | 0.04 | USA (18.3) |
| SR3_G8C     | 656     | 0.06     | 205   | 0.04 | 0   | 0 | 0 | 0 | 0 | 0 | USA (27.7) |
| SR3_T9C     | 465     | 0.04     | 1     | 0 | 0 | 0 | 0 | 0 | 0 | 0 | India (61.7) |
| SP2_G2T     | 467     | 0.04     | 14    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | The Netherland (45.8) |
| SP2_G20T    | 2818    | 0.26     | 308   | 0.07 | 13  | 0.11 | 8 | 0.07 | 1 | 0.04 | USA (18.3) |
| SP2_T22C    | 965     | 0.58     | 13    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | USA (19.0) |

* Sequence data (by 9 May 2012) were obtained from GISAID (https://www.gisaid.org/) filtering “complete”, “high coverage”, “low coverage excl” and “human”. The data of four VOCs were obtained following the classification by GISAID. Alignments were generated on MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/add_fragments.html?frommanual) based on Wuhan-Hu-1 sequence (MN908947.3. The mismatches in primer/probe sequences of N2 and S2 sets were searched using Sequencher software (ver5.4.6).
