Viral transmission and evolution dynamics of SARS-CoV-2 in shipboard quarantine

Ting-Yu Yeh* & Gregory P Contreras*

Objective To examine transmission and evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in shipboard quarantine of the Diamond Princess cruise ship.

Methods We obtained the full SARS-CoV-2 genome sequences of 28 samples from the Global Initiative on Sharing All Influenza Data database. The samples were collected between 10 and 25 February 2020 and came for individuals who had been tested for SARS-CoV-2 during the quarantine on the cruise ship. These samples were later sequenced in either Japan or the United States of America. We analysed evolution dynamics of SARS-CoV-2 using computational tools of phylogenetics, natural selection pressure and genetic linkage.

Findings The SARS-CoV-2 outbreak in the cruise most likely originated from either a single person infected with a virus variant identical to the WIV04 isolates, or simultaneously with another primary case infected with a virus containing the 11083G > T mutation. We identified a total of 24 new viral mutations across 64.2% (18/28) of samples, and the virus evolved into at least five subgroups. Increased positive selection of SARS-CoV-2 were statistically significant during the quarantine (Tajima’s $D: -2.03, P < 0.01$; Fu and Li’s $D: -2.66, P < 0.01$; and Zeng’s $E: -2.37, P < 0.01$). Linkage disequilibrium analysis confirmed that ribonucleic acid (RNA) recombination with the 11083G > T mutation also contributed to the increase of mutations among the viral progeny.

Conclusion The findings indicate that the 11083G > T mutation of SARS-CoV-2 spread during shipboard quarantine and arose through de novo RNA recombination under positive selection pressure.

Introduction

On 31 December 2019, Chinese authorities alerted the World Health Organization (WHO) of an outbreak of a novel coronavirus causing severe illness with pneumonia-like symptoms. The virus was later named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).1 WHO declared the outbreak a Public Health Emergency of International Concern on 30 January 2020 and as of 17 December 2020, more than 76 million confirmed cases had been reported and 1.6 million people had died.2

SARS-CoV-2 contains a single positive stranded RNA (ribonucleic acid) of 30 kilobases, which encodes for 10 genes,3 and like other RNA viruses, evolves through random nucleotide substitutions. Early sequencing of samples from coronavirus disease 2019 (COVID-19) patients on 30 January 2020 showed only six mutations compared with the first isolate WIV04, and hence suggested that the virus had a short infection history in humans.4 By April 2020, researchers had found more than 131 mutations in SARS-CoV-2 across the 103 sequenced viral genomes and they estimated that the virus accumulates about one to two mutations per month.5

On 25 January 2020, a passenger disembarked the Diamond Princess cruise ship in Hong Kong Special Administrative Region, China, and on 1 February, the passenger tested positive for SARS-CoV-2. When the ship docked in Yokohama, Japan, on February 3, the 3711 cruise passengers and crew members had to quarantine on the ship. As of 21 February 2020, 712 (19.2%) of these individuals had tested positive for SARS-CoV-2, of which 331 (46.5%) were asymptomatic at the time of testing. Among the 381 symptomatic individuals, 37 (9.7%) required intensive care and nine (2.4%) died.6,7

The shipboard quarantine provided a closed environment to observe the SARS-CoV-2 transmission and adaptation independently from other infectious resources.6,9 This environment presents an ideal static population, with little interfering noise, to measure the viral phylodynamics from the COVID-19 outbreak. We therefore decided to use this opportunity to study de novo evolution of SARS-CoV-2 in a closed population.

Methods

Data resources

Viral sequences and sequencing methods are available in the Global Initiative on Sharing All Influenza Data (GISAID)9 and GenBank databases.9 From these databases, we downloaded sequences and annotations of the isolates from the cruise ship, as well as the reference genomes of SARS-CoV-2 isolates PBCAMS-WH-04 (accession no. MT019532), WIV04 (MN996528), Hu-1 (NC045512) and WHU01 (MN98868).4

Statistical and phylogenetics analyses

We aligned FASTA files of viral sequences using MAFFT 7 software (Kazutaka Katoh, Research Institute for Microbial Diseases, Osaka, Japan).10,11 To analyse phylogenetic relationships between viral sequences, we used the neighbour-joining method and Jukes–Cantor substitution model with setting bootstrap resampling number as five. We generated the rectangular phylogenetic tree using Archaeopteryx with Java plug-in of MAFFT 7.12 For radial phylogenetic tree, we first exported the tree file as Newick format by MAFFT; the FigTree software (version 1.4.2, Andrew Rambaut, University of Edinburgh, Edinburgh, United Kingdom of Great Britain and Northern Ireland) was used to transform and display the cladogram.13 To illustrate the viral
Table 1. Characteristics of SARS-CoV-2 genomes from samples taken from people on shipboard quarantine, February 10 to February 25, 2020

| GISAID name       | Accession no.         | Specimen source     | Collection date | Genomic change | Type of mutation | Gene/protein | Amino acid change |
|-------------------|-----------------------|---------------------|-----------------|---------------|------------------|--------------|-------------------|
| hCoV-19/USA/CruiseA-1/2020 | EPI_ISL_413606        | Nasopharyngeal swab | 17 Feb          | 3099C > T     | Missense         | Orf1ab/NSP3   | T945I             |
| hCoV-19/USA/CruiseA-2/2020 | EPI_ISL_413607        | Nasopharyngeal swab | 18 Feb          | 28378G > T    | Synonymous       | N             | NA                |
| hCoV-19/USA/CruiseA-3/2020 | EPI_ISL_413608        | Nasopharyngeal swab | 18 Feb          | 28409C > T    | Missense         | N             | P465              |
| hCoV-19/USA/CruiseA-4/2020 | EPI_ISL_413609        | Nasopharyngeal swab | 21 Feb          | 29736C > T    | Non-coding       | 3′-UTR        | NA                |
| hCoV-19/USA/CruiseA-5/2020 | EPI_ISL_413610        | Oropharyngeal swab  | 21 Feb          | No change     | Synonymous       | Orf1ab/NSP2   | NA                |
| hCoV-19/USA/CruiseA-6/2020 | EPI_ISL_413611        | Nasopharyngeal swab | 21 Feb          | 11410G > A    | Missense         | Orf1ab/NSP6   | NA                |
| hCoV-19/USA/CruiseA-7/2020 | EPI_ISL_413612        | Nasopharyngeal swab | 17 Feb          | 26326C > T    | Synonymous       | E             | NA                |
| hCoV-19/USA/CruiseA-8/2020 | EPI_ISL_413613        | Nasopharyngeal swab | 17 Feb          | 9474C > T     | Missense         | Orf1ab/NSP4   | A3070V            |
| hCoV-19/USA/CruiseA-9/2020 | EPI_ISL_413614        | Nasopharyngeal swab | 17 Feb          | No change     | NA               | NA            | NA                |
| hCoV-19/USA/CruiseA-10/2020| EPI_ISL_413615        | Nasopharyngeal swab | 17 Feb          | 3259G > T     | Missense         | Orf1ab/NSP3   | Q998H             |
| hCoV-19/USA/CruiseA-11/2020| EPI_ISL_413616        | Nasopharyngeal swab | 17 Feb          | 10036C > T    | Non-coding       | 3′-UTR        | NA                |
| hCoV-19/USA/CruiseA-12/2020| EPI_ISL_413617        | Oropharyngeal swab  | 20 Feb          | No change     | Synonymous       | E             | NA                |
| hCoV-19/USA/CruiseA-13/2020| EPI_ISL_413618        | Nasopharyngeal swab | 20 Feb          | 6636C > T     | Missense         | Orf1ab/NSP3   | T212H             |
| hCoV-19/USA/CruiseA-14/2020| EPI_ISL_413619        | Oropharyngeal swab  | 25 Feb          | 11750C > T    | Missense         | Orf1ab/NSP6   | L3829F            |
| hCoV-19/USA/CruiseA-15/2020| EPI_ISL_413620        | Nasopharyngeal swab | 18 Feb          | 11956C > T    | Synonymous       | Orf1ab/NSP7   | NA                |
| hCoV-19/USA/CruiseA-16/2020| EPI_ISL_413621        | Nasopharyngeal swab | 18 Feb          | No change     | NA               | NA            | NA                |
| hCoV-19/USA/CruiseA-17/2020| EPI_ISL_413622        | Nasopharyngeal swab | 24 Feb          | 5845A > T     | Missense         | Orf1ab/NSP3   | K1860N            |
| hCoV-19/USA/CruiseA-18/2020| EPI_ISL_413623        | Nasopharyngeal swab | 24 Feb          | 508_522del    | Deletion         | Orf1ab/NSP1   | Deletion 82–86    |
| hCoV-19/USA/CruiseA-19/2020| EPI_ISL_414479        | Nasopharyngeal swab | 18 Feb          | No change     | NA               | S             | F157L             |
| hCoV-19/USA/CruiseA-20/2020| EPI_ISL_414480        | Oropharyngeal swab  | 21 Feb          | 254C > T      | Non-coding       | 3′-UTR        | NA                |
| hCoV-19/USA/CruiseA-21/2020| EPI_ISL_414481        | Nasopharyngeal swab | 21 Feb          | No change     | NA               | NA            | NA                |

(continues . . .)
| Specimen source       | Collection date | Type of mutation | Gene/protein | Accession no.                                      | GISAID name                                      |
|----------------------|-----------------|------------------|--------------|----------------------------------------------------|-------------------------------------------------|
| Nasopharyngeal swab  | 18 Feb          | Missense         | Orf1ab       | EPJL_414482                                       | hCoV-19/USA/CruiseA-23/2020                     |
|                      |                 |                  |              | EPJL_414483                                       | hCoV-19/USA/CruiseA-24/2020                     |
|                      |                 | Missense         | Orf1ab       | EPJL_414484                                       | hCoV-19/USA/CruiseA-25/2020                     |
|                      |                 |                  | Orf1ab       | EPJL_414485                                       | hCoV-19/USA/CruiseA-26/2020                     |
|                      |                 | Missense         | Orf1ab       | EPJL_414486                                       | hCoV-19/Japan/Hu_DP_Kng_19–027/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414487                                       | hCoV-19/Japan/Hu_DP_Kng_19–020/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414488                                       | hCoV-19/Japan/Hu_DP_Kng_19–029/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414489                                       | hCoV-19/Japan/Hu_DP_Kng_19–037/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414490                                       | hCoV-19/Japan/Hu_DP_Kng_19–047/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414491                                       | hCoV-19/Japan/Hu_DP_Kng_19–057/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414492                                       | hCoV-19/Japan/Hu_DP_Kng_19–067/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414493                                       | hCoV-19/Japan/Hu_DP_Kng_19–077/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414494                                       | hCoV-19/Japan/Hu_DP_Kng_19–087/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414495                                       | hCoV-19/Japan/Hu_DP_Kng_19–097/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414496                                       | hCoV-19/Japan/Hu_DP_Kng_19–107/2020             |
|                      |                 | Missense         | Orf1ab       | EPJL_414497                                       | hCoV-19/Japan/Hu_DP_Kng_19–117/2020             |

To determine whether the viral genome undergoes neutral or non-neutral evolution, that is, genetic variations of viral genomes are due to randomly genetic drift or under natural selection pressure, we used MEGA7 software to calculate Tajima’s D test of neutrality. This method compares the numbers of mutations per site with the nucleotide diversity (the mean pairwise difference between sequences). To compare the number of derived singleton site mutations – that is, single base mutations that occur only once in a given population – and the mean pairwise difference between sequences, we calculated Fu and Li’s D, using online PopS website calculator (Shi-Yi Chen, Sichuan Agricultural University, Chengdu, China). We used the same calculator for calculating Zeng’s E, which measures changes in high-frequency variants. To calculate the P values for Tajima’s D, Fu and Li’s D and Zeng’s E values, we used DnaSP 6 software (Julio Rozas, University of Barcelona, Barcelona, Spain).

To investigate the linkage disequilibrium, that is the non-random assortment of alleles at different loci, of SARS-CoV-2 genomes, we first converted 148 SARS-CoV-2 genomic sequences using SNP tools plug-in in Excel (Microsoft, Redmond, United States of America, USA) to create a baseline. We downloaded these sequences from GISAID. Using HaploView software, version 4.1 (Broad Institute, Cambridge, USA), we measured and plotted the normalized values (D’) of the coefficient of linkage disequilibrium (D). We obtained D’ by dividing D with D_{max} where D_{max} is the theoretical maximum difference between the observed and expected haplotype frequencies. We also calculated the log of the odds of there being a disequilibrium between two loci and the squared coefficient of correlation (r²) using the same software. In the absence of evolutionary forces or natural selection, the D’ converges to zero along the time axis at a rate depending on the magnitude of the recombination rate between the two loci. We used the χ² test to examine if the obtained linkage disequilibrium was statistically significant. To detect positive RNA recombination, we plotted 95% confidence intervals. We compared the genomes to SARS-CoV-2 sequence to WIV04 (accession no. KT444582) as an outgroup virus.

Selection pressure

To determine whether the viral genome undergoes neutral or non-neutral evolution, that is, genetic variations of viral genomes are due to randomly genetic drift or under natural selection pressure, we used MEGA7 software to calculate Tajima’s D test of neutrality. This method compares the number of mutations per site with the nucleotide diversity (the mean pairwise difference between sequences). To compare the number of derived singleton site mutations – that is, single base mutations that occur only once in a given population – and the mean pairwise difference between sequences, we calculated Fu and Li’s D, using online PopS website calculator (Shi-Yi Chen, Sichuan Agricultural University, Chengdu, China). We used the same calculator for calculating Zeng’s E, which measures changes in high-frequency variants. To calculate the P values for Tajima’s D, Fu and Li’s D and Zeng’s E values, we used DnaSP 6 software (Julio Rozas, University of Barcelona, Barcelona, Spain).

To investigate the linkage disequilibrium, that is the non-random assortment of alleles at different loci, of SARS-CoV-2 genomes, we first converted 148 SARS-CoV-2 genomic sequences using SNP tools plug-in in Excel (Microsoft, Redmond, United States of America, USA) to create a baseline. We downloaded these sequences from GISAID. Using HaploView software, version 4.1 (Broad Institute, Cambridge, USA), we measured and plotted the normalized values (D’) of the coefficient of linkage disequilibrium (D). We obtained D’ by dividing D with D_{max} where D_{max} is the theoretical maximum difference between the observed and expected haplotype frequencies. We also calculated the log of the odds of there being a disequilibrium between two loci and the squared coefficient of correlation (r²) using the same software. In the absence of evolutionary forces or natural selection, the D’ converges to zero along the time axis at a rate depending on the magnitude of the recombination rate between the two loci. We used the χ² test to examine if the obtained linkage disequilibrium was statistically significant. To detect positive RNA recombination, we plotted 95% confidence intervals. We compared the genomes to SARS-CoV-2 sequence to WIV04 (accession no. KT444582) as an outgroup virus.
Results

Viral variants

A total of 28 specimens with viral sequences were available for this analysis, including 25 samples from the United States and three samples from Japan. Table 1 lists the characteristics of the viral sequences. Genetic variations of viral sequences were present in 71.4% (20/28) of samples. The sequences of eight samples were completely identical to the Wuhan isolates PBCAMS-WH-04, WIV04, Hu-1 and WHU01. A total of 24 new substitution mutations were identified in 18 samples (Table 1 and Fig. 1).

Two mutations had previously been identified in other SARS-CoV-2 isolates. First, the deletion in hCoV-19/USA/CruiseA-18/2020 has been detected in USA-CA6 (MT044258). However, the possibility that hCoV-19/USA/CruiseA-18/2020 contributed to a new mutation in the cruise samples can be dismissed because this deletion was absent in the other samples. Second, the nonsynonymous mutation 11083G > T was found in USA-AZ1 (EPI_ISL_406223) on 22 January. Later, this mutation was also present in WA3-UW1 (EPI_ISL_413025), NY-NYUMC1 (EPI_ISL_414639) and UPHL-01 (EPI_ISL_415539U).

Viral origin and transmission

Fig. 1 shows the phylogenetic trees of the 28 samples without introducing the outgroup virus. Half (14/28) of the samples contained a virus variant that had evolved, that is, had more than two mutations, during the shipboard quarantine. We identified five subgroups after rooting the phylogenetic tree with an outgroup virus sequence of SARS-WIV16: (i) 3099C > T and 28378G > T (two samples); (ii) 11083G > T (three samples); (iii) 11410G > A and 26326C > T (two samples); (iv) 29635C > T (four samples); and (v) 29736G > T (three samples; Fig. 2). Due to 11083G > T mutation, the clustering of taxa on viral phylogenies was obvious with spatially structured host population between all these subgroups (50%; 14/28; Fig. 2).

Whether the single 11083G > T substitution spontaneously occurred during the quarantine or the patients had been infected with a viral variant containing this mutation before boarding the ship is unclear. Nevertheless, all of the viral sequences were more similar to the WIV04 sequence than with other 143 SARS-CoV-2 isolates in the GISAID database (data repository). This result suggests that the 24 new mutations identified were generated de novo on the ship rather than deriving from multiple geographical origins.

The analysis revealed two possibilities of the viral origin: either the virus (except hCoV-19/USA/CruiseA-18/2020) originated from a single primary case with the WIV04 sequence and all substitution mutations occurred...
during the quarantine; or there were two simultaneously primary cases, one identical to the WIV04 sequence and one containing the 11083G > T substitution.

**Natural selection of mutations**

Four variants had three mutations in their genome, one variant had four mutations and two variants had six mutations (Fig. 1 and Table 1). To test the hypothesis that the virus mutation evolved under selection pressure as opposed to neutral evolution (random) onboard the cruise ship, we calculated Tajima’s D to test neutrality of DNA polymorphisms.\textsuperscript{15,20} The Tajima’s D value in the cruise was −2.03 (P < 0.01) compared with 6.75 among 143 full-length genomic sequences of SARS-CoV-2 isolates sequenced between 10 January and 13 March 2020. This result indicates that, while other SARS-CoV-2 isolates faced balancing selection (due to a positive Tajima’s D), the virus spreading on the cruise had evolved under two possible but not exclusive forces generated during the quarantine process: purifying or positive selection; and population growth of the new virus variant among infected patients after a recent bottleneck as a force. We conclude that SARS-CoV-2 viral evolution was positively correlated to the increase of the selection pressure during the shipboard quarantine.

**RNA recombination**

Seven samples contain the 11083G > T mutation although they belong to different subgroups (Fig. 1). By assuming the substitution rate of 0.92 × 10\textsuperscript{−3}/site/year,\textsuperscript{26} it is unlikely that the virus variants of different subgroups all generated the same spontaneous mutation at the G11083 site into the nucleotide T within three weeks. One hypothesis is that RNA recombination occurred in these cases to gain the 11083G > T mutation.

To determine whether four variants from different subgroups (samples /USA/CruiseA-23/, /USA/CruiseA-24/, /Japan/TK/20–31–3/2020 and /Japan/Hu_DK_Kng19–027/2020) obtained the G11083 > T mutation via RNA recombination, we analysed the patterns of linkage disequilibrium between variants with minor alleles of two SARS-CoV-2 variants (available in the data repository).\textsuperscript{24} Table 2 shows five pairs of mutations among all analysed samples with high log of the odds of there being a disequilibrium between two loci (> 3.0) and their r\textsuperscript{2} values. The linkages between double mutations in variants /USA/CruiseA-1/ and /USA/CruiseA-24/, /USA/CruiseA-6/ and /USA/CruiseA-7/, and /USA/CruiseA-21/ and /USA/CruiseA-23/ were statistically significant (P < 0.001). Furthermore, there was indication of positive RNA recombination. The upper 95% confidence bound of D’ for C254/ G11083 was 0.86, for G29736/G11083 was 0.73, for G29751/G11083 was 0.86, for C3099/G11083 was 0.86, for G28378/ G11083 was 0.86 and for C29635/G11083 was 0.80 (Fig. 3). This result supports the hypothesis that 11083G > T mutation had been gained via RNA recombination in these four variants.

We detected the mutations 11083G > T and 26326 C > T, present in the UPHL-01 sequence, in the two variants: /USA/CruiseA-6/ and /USA/CruiseA-7/. The upper 95% confidence bounds on D’ of C26326/G11083 was 0.73 (Fig. 3),
sustaining that a recent RNA recombination event may have also occurred in UPHL-01.

Fig. 4 shows each haplotype in the two identified blocks and their population frequencies and connections between blocks. The value of multi-allelic $D'$ was 0.65, which represented the level of recombination between the two blocks. These data provide evidence that RNA recombination contributed to SARS-CoV-2 mutations during shipboard quarantine.

**Protein mutations**

The viral populations in the cruise also generated 11 new variants in the viral proteins, including proteins in the gene ORF1ab (nonstructural protein NSP3, NSP4, NSP6), and in the viral structural spike and nucleocapsid proteins (Table 1). We found five missense variants in NSP3: T945I; Q998H; P1158S in the macro domain; K1860N in papain-like protease; and T2124I in the group 2 marker domain. In the other NSPs, we detected two mutations: A3070V in NSP4 and L3829F in NSP6. For the spike protein, two mutations, F157L and G181V, were identified. For the nucleocapsid phosphoprotein, we found mutations in the RNA binding domain (P46S) and the arginine-serine domain (G215S).

Only variant /USA/CruiseA-14/ had two mutations, T2124I and L3829F in NSP3 and NSP6, respectively (Table 1).

**3'-UTR mutations**

In three samples, we found two mutations, 29736G > T and 29751G > T in the stem loop-II motif (Fig. 5). We used the published three-dimensional crystal structure of SARS-CoV stem loop-II motif RNA to map the nucleotides G29736 and G29751. We found that these nucleotides are equal to the nucleotides G13 and G28 in the SARS-CoV stem loop-II motif (Fig. 5). In SARS-CoV, G13 (G29736 in SARS-CoV-2) forms a base triple with A38 and C39 in a seven-nucleotide asymmetric bubble, while G28 (G29751 in SARS-CoV-2) participates in formation of an essential RNA base quartet composed of two G–C pairs (G19, C20, G28, C31; Fig. 5).

**Discussion**

Viral phylogenetics is a useful tool to study epidemiological and evolutionary processes, such as epidemic spread and spatiotemporal dynamics including metapopulation dynamics, zoonotic transmission, tissue tropism and antigenic drift. Here we report the viral phylodynamics of SARS-CoV-2 from patients in a shipboard quarantine for three weeks in February 2020. The transmission started from either one or two primary cases with WIV04 sequence and/or 11083G > T mutation, then quickly separated into at least five subgroups based on new mutations. Increased positive selection as well as RNA recombination of SARS-CoV-2 were evident during the quarantine. These results should be considered in formulation of future management protocols with respect to a SARS-CoV-2 outbreak.

### Table 2. Linkage disequilibrium of SARS-CoV-2 mutations with high log of the odds and the $r^2$ values

| Pair                  | Log of the odds | $r^2$ |
|-----------------------|-----------------|-------|
| C3099/G28378          | 4.60            | 1.00  |
| G11410/C26326         | 3.77            | 0.66  |
| G29736/29751          | 3.77            | 0.66  |
| C254/G29736           | 3.77            | 0.66  |
| C254/G29751           | 4.60            | 1.00  |

**Fig. 3. Haplotype block organization of SARS-CoV-2 mutations in samples from the cruise ship Diamond Princess, 2020**

**Fig. 4. Haplotype frequencies of SARS-CoV-2 mutations in the cruise ship Diamond Princess, 2020**

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

Notes: Linkage disequilibrium plot of HaploView to display the confidence bounds colour scheme. Each box represents a pair of mutations. The solid spines of strong linkage disequilibrium running from one marker to another along the legs of the triangle in the linkage disequilibrium chart determine the haplotype block. We defined strong evidence of recombination if pairs for which the upper confidence bound of the coefficient of linkage disequilibrium is less than 0.9 (white squares). We did not detect any strong linkage.

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

Notes: Numbers next to each haplotype block are haplotype frequencies. The joining lines represent combined haplotypes. In the crossing areas between haplotype blocks, a value of multi-allelic $D’$ that is the normalized value of the coefficient of linkage disequilibrium, is shown to represent the level of recombination between blocks.
in any relatively close quarters, such as shipboards, submarines, dormitories, prisons and hospitals.

While the quarantine averted a lot of infections on the shipboard, the phylogenetics analysis showed that viral transmission and RNA recombination occurred between the five identified subgroups. Our data fit in the coalescent model, which uses the diversity of viral genome, the viral evolutionary rate and the estimated time of infection to determine the number of viral genotypes present in the initial infected population. However, we cannot rule out that evolutionary processes, such as the transmission bottleneck that determines how much of the viral diversity generated in one host passes to another during transmission, also shaped the viral phylogenies. While spatial structure is the most general virus population structure in phylodynamic analyses, SARS-CoV-2 evolution may also have been influenced by the characteristic of the host, such as age, race and risk behaviour. Because viral transmission can preferentially occur between patients sharing any of these attributes, the real reason(s) for viral transmission between virus variants require(s) further study. Furthermore, studies on whether quarantine in close quarters also promotes virus to rapidly gain more mutations via RNA recombination are needed.

Despite the small sample size in this study, our findings from computational and statistical analyses indicated that the selection pressure was not random. We assume that SARS-CoV-2 variants were at an initial stage of evolution rather than the fixation stage, since the COVID-19 outbreak had only started 8 weeks before the Diamond Princess incident. The tools we used to measure selection pressure of adaptive evolution were Tajima’s D, Zeng’s E and Fu and Li’s D tests, because of their appropriateness in relation to the sample characteristics. For Tajima’s D, its power is not affected by RNA recombination events, number of segregation sites and various timescales nor does its power is not affected by RNA recombination events, number of segregation sites and various timescales nor does it change with allele frequency. Zeng’s E test, because of the simplicity and robustness of this method. The reasons for this exclusion were: (i) the ratio is defined to describe the relative rate of selected versus neutral fixation events over long timescales, not transient polymorphisms over short timescales; and (ii) dN/dS value in a single population does not follow a monotonic function in propor-

Fig. 5. Mutations in the SARS-CoV-2 stem loop-II RNA motif

| Virus strain          | Nucleotide site G13 | Sequence site G28 | Nucleotide site G28 |
|-----------------------|---------------------|------------------|---------------------|
| hCoV-19/USA/CruiseA-2/2020 | UUCACGAU          | GCC AG            | CG UGU ACA          |
| hCoV-19/USA/CruiseA-21/2020  | UUCACGAU        | GCC AG            | CG UGU ACA          |
| hCoV-19/USA/CruiseA-23/2020  | UUCACGAU        | GCC AG            | CG UGU ACA          |
| Other hCoV-19/USA/CruiseA isolates | UUCACGAU    | GCC AG            | CG UGU ACA          |
| WIV04                  | UUCACGAU          | GCC AG            | CG UGU ACA          |
| SARS-CoV               | UUCACGAU          | GCC AG            | CG UGU ACA          |

SARS-CoV: severe acute respiratory syndrome coronavirus.

Notes: We compared hCoV-19/USA/CruiseA-2/2020 (accession no. EPI_ISL_413607), hCoV-19/USA/CruiseA-21/2020 (EPI_ISL_414480), hCoV-19/USA/CruiseA-23/2020 (EPI_ISL_414482) and other cruise isolates with WIV04 (MN996528) and SARS-CoV (NC004718). Conventional RNA helical base-pairing are indicated in italics. The schematic representation of the stem loop-II RNA secondary structure of SARS-CoV virus shows nucleotide interactions. The digits 13 and 28 show the location of the mutations identified in some of the cruise ship isolates and digit 19 represents the location of stem loop-II mutation found in Australia.
The mutation appeared independently of cruise ship variants is unknown. Because this mutation was also later detected in other variants, in addition to UPHL-01, future studies should further investigate whether 11083G > T may increase the fitness of the carrier. Other studies have suggested the 11083G > T could be a beneficial mutation linked to asymptomatic infection.27,36

Of the 24 mutations we identified, 11 mutations led to amino acid substitutions and two mutations occurred in the stem loop-II motif in the 3'UTR region. This motif is a well-conserved RNA motif in more than 30 coronaviruses.29,36 We have also reported the unique 29742G > A or 29742G > U substitutions in stem loop-II motif RNA in SARS-CoV-2 isolates in Australia (Fig. 5),28 reinforcing the idea that stem loop-II motif is a hotspot for mutations in SARS-CoV-2 rather than a conserved RNA domain.27

Mutations in this motif may disrupt RNA structure and thereby alter the viral viability or infectivity. Whether the mutations identified here, both in viral proteins and regulatory RNA regions, may enhance adaptation or attenuate virus replication or virulence9 requires further investigation.

We acknowledge there may be limitations in collection of patients’ samples and size in extending our findings to represent the entire SARS-CoV-2-positive population (712 people, from January 20 to March 8, 2020) in the Diamond Princess cruise.6 If possible, further investigation with more viral genome sequences is required to understand the detailed evolutionary lineage of transmission and whether there is difference of viral mutations between symptomatic and asymptomatic individuals. ■

Competing interests: None declared.
研究结果显示，SARS-CoV-2 病毒发生 11083G > T 突变传播的原因在于在船舶检疫隔离期间正性选择压力下 RNA 重组的新发突变。

### Résumé

Transmission virale et dynamique d'évolution du SARS-CoV-2 lors d'une quarantaine en mer

**Objectif** Étudier la transmission et l'évolution du coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) lors d'une quarantaine à bord du navire de croisière Diamond Princess.

**Méthodes** Nous avons obtenu l'ensemble des séquences génétiques du SARS-CoV-2 de 28 échantillons issus de la base de données de l'Initiative mondiale de partage des données sur la grippe aviaire (GISAID). Ces échantillons ont été collectés entre le 10 et le 25 février 2020 et prélevés sur des individus soumis à un test de dépistage du SARS-CoV-2 durant la quarantaine à bord du navire. Ils ont ensuite été séquencés soit au Japon, soit aux États-Unis d’Amérique. Nous avons analysé la dynamique d'évolution du SARS-CoV-2 à l'aide d'outils informatiques de phylogénétique, de pression sélective naturelle et de liaison génétique.

**Résultats** L'épidémie de SARS-CoV-2 à bord provient vraisemblablement d'une seule personne infectée par un variant identique aux isolats WIV04, ou simultanément d'un autre cas primaire d'infection par un virus contenant la mutation 11083G > T. Nous avons détecté au total 24 nouvelles mutations virales dans 64,2% (18/28) des échantillons, et le virus a évolué en cinq sous-groupes minimum. La sélection positive accrue de SARS-CoV-2 s'est révélée statistiquement significative durant la quarantaine (D de Tajima: −2,03, P < 0,01; D de Fu et Li: −2,66, P < 0,01; et E de Zeng: −2,37, P < 0,01). L'analyse du déséquilibre de liaison a confirmé que la recombinaison de l'acide ribonucléique (ARN) avec la mutation 11083G > T avait également contribué à la hausse des mutations dans la lignée virale.

**Conclusion** Ces résultats indiquent que la mutation 11083G > T du SARS-CoV-2 s'est répandue durant la quarantaine à bord du navire, et par le biais d'une recombinaison de novo de l'ARN sous une pression sélective positive.

### Resumen

Transmisión viral y dinámica de la evolución del SARS-CoV-2 durante la cuarentena en cruceros

**Objetivo** Examinar la transmisión y la evolución del coronavirus del síndrome respiratorio agudo grave 2 (SARS-CoV-2) en la cuarentena del crucero Diamond Princess.

**Métodos** Obtuvimos las secuencias completas del genoma del SARS-CoV-2 de 28 muestras de la base de datos de la Global Initiative on Sharing All Influenza Data. Las muestras se recogieron entre el 10 y el 25 de febrero de 2020 y procedían de individuos a los que se les había hecho la prueba del SARS-CoV-2 durante la cuarentena en el crucero. Estas muestras se secuenciaron posteriormente en Japón o en los Estados Unidos de América. Se analizó la dinámica evolutiva del SARS-CoV-2 utilizando herramientas computacionales de filogenética, presión de selección natural y enlace genético.

**Resultados** Lo más probable es que el brote de SARS-CoV-2 en el crucero se haya originado a partir de una sola persona infectada con una variante del virus idéntica a las cepas WIV04, o bien simultáneamente con otro caso primario infectado con un virus que contenía la mutación 11083G > T. Los antecedentes de 24 nuevas mutaciones virales en 64,2% (18/28) de los muestras, y el virus evolucionó como mínimo en cinco subgrupos. El aumento de la presión selectiva del SARS-CoV-2 fue estadísticamente significativo durante la cuarentena (D de Tajima: −2,03, P < 0,01; D de Fu y Li: −2,66, P < 0,01; y E de Zeng: −2,37, P < 0,01). El análisis de desequilibrio de línea de árbol confirmó que la recombinación del ácido ribonucleico (ARN) con la mutación 11083G > T también contribuyó al aumento de las mutaciones entre la progenie viral.

**Conclusión** Los resultados indican que la mutación 11083G > T del SARS-CoV-2 se propagó durante la cuarentena en el crucero y surgió por recombinación de novo del ARN bajo presión de selección positiva.
