Archives of Virology (2021) 166:1735–1739
https://doi.org/10.1007/s00705-021-05029-7

BRIEF REPORT

Diverse SARS-CoV-2 variants preceded the initial COVID-19 outbreak in Croatia

Filip Rokić[1] · Lovro Trgovec-Greif[1] · Neven Sučić[2] · Noa Čemeljić[3,6] · Đurđica Cekinović Grbeša[4,5] · Željko Svedružić[6] · Tomislav Rukavina[2,4] · Oliver Vugrek[1] · Igor Jurak[6]

Received: 20 October 2020 / Accepted: 21 January 2021 / Published online: 24 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

We developed a next-generation SARS-CoV-2 sequencing platform and obtained the first SARS-CoV-2 sequences from patients in Croatia at the beginning of the COVID-19 outbreak in the spring of 2020. Integrating the sequencing and the epidemiological data, we show that patients were infected with different SARS-CoV-2 variants belonging to different clades (mostly G and GH). This result confirms that there was widespread virus transmission early in 2020. Interestingly, we identified a unique mutation resulting in a V13I substitution in Nsp5A, the main viral protease, in a patient who had not received antiviral therapy.

The COVID-19 pandemic has encircled the world, taking an unprecedented toll in human lives and economic losses. In many European countries, SARS-CoV-2 was introduced relatively soon after the initial outbreak and spread uncontrollably during January and February 2020 [3, 4, 7, 14, 25]. This resulted in a massive influx of patients into hospitals, in some cases overwhelming the national capacities, leading to strict epidemiological measures and restrictions [9, 24]. The first cases of COVID-19 in Croatia were identified on February 25, 2020, leading to an outbreak that peaked during April [12]. Here, based on sequencing and epidemiological data, we report genetic evidence for multifocal introduction of different SARS-CoV-2 variants during the early COVID-19 epidemic in the spring of 2020 in Croatia.

Handling Editor: Zhenhai Chen.

✉ Oliver Vugrek
ovugrek@irb.hr

✉ Igor Jurak
igor.jurak@biotech.uniri.hr

1 Laboratory for Advanced Genomics, Ruđer Bošković Institute, Zagreb, Croatia

2 Teaching Institute for Public Health, Rijeka, Croatia

3 Master's Programme in Biomedicine, Karolinska Institutet, Stockholm, Sweden

4 Faculty of Medicine in Rijeka, University of Rijeka, Rijeka, Croatia

5 Department of Infectious Diseases, University Hospital Rijeka, Rijeka, Croatia

6 Department of Biotechnology, University of Rijeka, Rijeka, Croatia

Samples and SARS-CoV-2 whole-genome sequencing

To investigate possible routes of SARS-CoV-2 entry into Primorje-Gorski Kotar County, Croatia, we sequenced 21 RNAs derived from nasopharyngeal/throat swabs collected from symptomatic patients and their suspected contacts during the early COVID-19 epidemic in Croatia (between March 22 and April 10, 2020). Prior to sequencing, SARS-CoV-2 was confirmed in all samples by RT-qPCR (GeneFinder™ COVID-19 PLUS RealAmp Kit, GenDx), and basic epidemiologically relevant information was collected, including age, gender, contacts, and recent travel history. Of note, the vast majority of samples were positive for all three genes tested (E, N, and RdRp) (not shown). To sequence the genome of the virus, we designed an in-house next-generation sequencing panel (NGS) for the Illumina platform covering the complete SARS-CoV-2 genome by 173 amplicons (Supplemental Table 1). Briefly, total RNA was extracted from the swab samples and used to generate cDNA using a ProtoScript II First Strand cDNA Synthesis Kit (NEB). Each cDNA sample was used as a template in four separate multiplex PCR reactions with the in-house-designed SARS-CoV-2 specific primer pools covering the entire viral

genome (Supplementary Table 1). PCR reaction products were purified using AMPure XP beads (Beckman Coulter), quantified, and used as input for NGS library preparation (NEBNext Ultra II, NEB) according to the manufacturer's protocol. Finally, libraries were sequenced using MiniSeq (Illumina), and the resulting sequences were aligned to the reference genome sequence NC_45512.2.

For seven samples, we obtained high-quality nearly complete genome sequences with gaps of 350-850 nucleotides and an average sequence coverage between 6000 and 12,000×. Ambiguous nucleotides and gaps were confirmed by direct Sanger sequencing, and the complete genome sequences were deposited in GISAID, the main open-source database of SARS-CoV-2 genome sequences (https://www.epicov.org) [6]. For the remaining 14 samples, we retrieved partial sequences, covering 50% to 84% of the viral genome, which were used for partial genomic and molecular epidemiology analysis.

Phylogenetic analysis, mutations and clades

Our sequence analysis shows that, compared to the SARS-CoV-2 reference genome (NC_045512.2; based on the early Wuhan isolate [28]), the first viruses sequenced in Croatia accumulated between five and nine high-confidence mutations (Supplementary Table 2). This result is similar to the number of mutations in viruses simultaneously sequenced in different countries, confirming that this virus has a relatively low mutation rate (i.e., about $0.80\text{-}2.38 \times 10^{-3}$ nucleotide substitution per site per year) (www.nextstrain.org.) [8, 19, 20, 26, 27, 31]. The majority of identified mutations were missense C-U transitions (60-80%; Supplementary Table 2). Interestingly, the strain with the largest number of mutations (LG, Fig. 1), which belongs to an unusual clade with low representation (i.e., currently classified as "other clade" at GISAID.org; more details below), accumulated more G-U (3x) and U-C (2x) mutations, and only one of its nine mutations was a C-U transition. At this point, we cannot explain this peculiarity. Furthermore, we found that the majority of the isolates had accumulated mutations in the 5' UTR, Nsp3, RdRp, spike, and ORF3a genes (Supplementary Table 2),

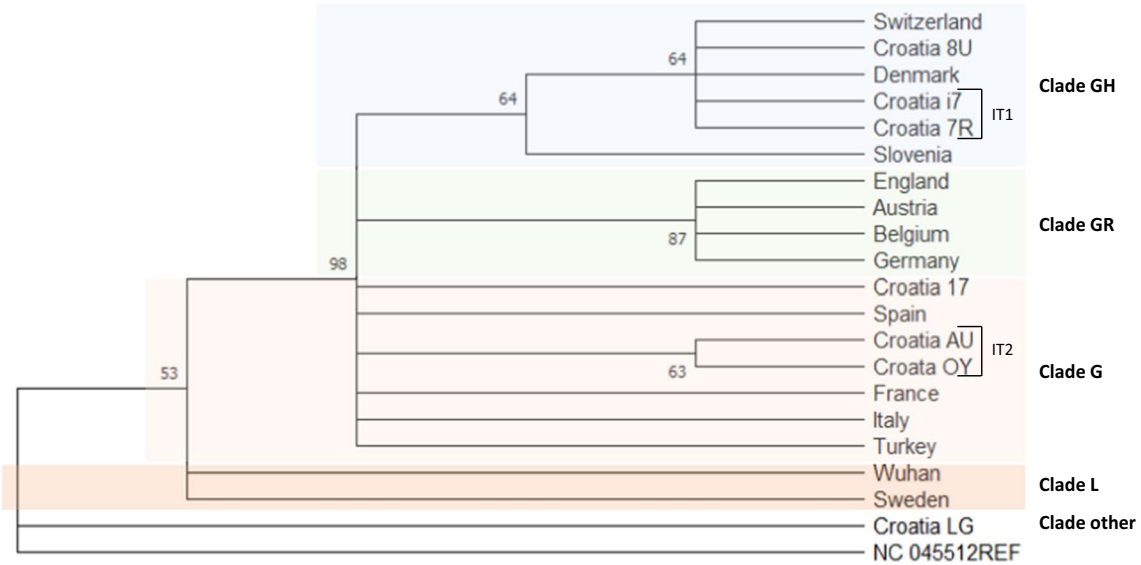

Fig. 1 Phylogenetic analysis of the first seven Croatian SARS-CoV2 sequences. Phylogeny reconstruction based on genome sequences of seven SARS-CoV-2 isolates from patients in Primorje-Gorski Kotar County between March 20 and April 10 2020 (Croatia 8U, i7, 7R, 17, AU, OY, and LG) and randomly selected genome sequences from 13 countries indicated in the figure (details in Supplementary Table 4) inferred by the maximum-likelihood method, the bootstrap test, and Tamura-Nei modeling, conducted in MEGA-X [16, 17]. All genomes were trimmed of highly variable regions, and nucleotides 55-29,836 (NC_045512.2) were aligned using MAFFT [19] or the MEGA-X-integrated MUSCLE multiple sequence alignment program [16, 17]. The number of bootstrap replications was 1000, with a cutoff value of 51% for the condensed tree. Bootstrap values are indicated at the branch nodes. The tree was rooted on the NC_045512.2 branch. The following clades are highlighted in different colors: clade G (substitutions C214T, C3037T, A23403G), GH (C214T, C3037T, A23403G, G25563T), GR (C241T, C3037T, A23403G, G28882A), and L (C241, C3037, A23403, C8782, G11083, G26144, T28144). NC 045512REF is the SARS-CoV-2 reference sequence (NCBI accession number NC_045512.2). IT1 (same patient) and IT2 (spouses) represent internal transmission 1 and 2, respectively.

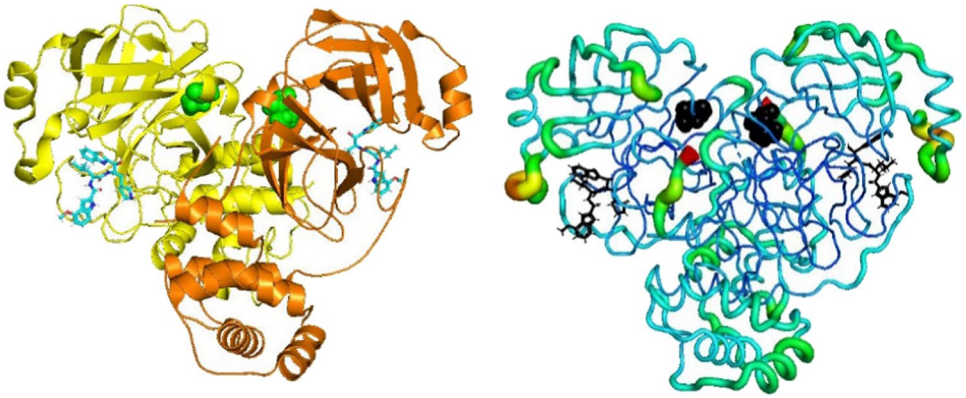

Fig. 2 Position of the V13I mutation in the structure of the SARS-Cov-2 Mpro protease with an inhibitor bound in the active site. The crystal structure of the protease (PDB:6XR3) [2] was used to show the position of the V13I mutation. A) The two subunits of the protease dimer are shown as ribbon models, in orange and yellow, respectively. The blue licorice model shows the drug GRL-024-20 bound in the active site of each monomer. The side chains of V13 at the position of the mutation are shown as CPK models, buried in the protein interior as part of the first helix of the protein. B) The crystallographic B-factor values are depicted using color, and the width of the ribbon represents the relative mobility in the protein structure; thin blue lines represent low mobility, yellow-green ribbons represent intermediate mobility, and thick red ribbons represent sites with high mobility. All models were prepared in VMD [10].

which has been also observed globally (nextstrain.org) [26, 27] and might indicate a strong fitness and/or adaptation advantage [15, 21]. The other observed mutations were within the Nsp2, Nsp5A, Nsp6, M, and N genes (Supplementary Table 2). Although our entire sample size is very small, the observed predominance of missense mutations in these genes (i.e., change of amino acid), compared to other genes in which fewer mutations were identified, might indicate the importance of these genes for adaptation of the virus to a new host or its transmissibility [reviewed in reference 5].

In our analysis, we utilized the GISAID dynamic phylogenetic framework, which was developed to facilitate genomic epidemiology efforts [22], and we found that the majority of the fully sequenced strains belonged to clade G (6/7; D614G in the S gene) (Fig. 1 and Supplementary Tables 2 and 3), three of which belonged to subclade GH. Moreover, analysis of the partially completed sequences available from the remaining 14 sequenced viruses showed that almost all of them (20/21) belonged to clade G. These results are in accordance with the overall distribution of identified strains in the world and in Europe (GISAID.org) at the time of sample collection [1, 11, 13, 23]. It is important to note that the frequency of the clades is changing rapidly and varies significantly over time and between different regions. For example, in January 2021, the dominant clade in Europe was GV, in South America it was GR, and in Asia it was GH (GISAID). The molecular basis for selection of variants and the clinical and epidemiological importance of novel mutations and clades are not well understood [18, 21], but it is clear that both national and global real-time surveillance based on virus genomics will be important for addressing these questions.

The dynamics of mutations

The epidemiological data that we collected allowed us to establish a direct transmission link for several patients (e.g., husband and wife, close relatives, etc.) and to address the dynamics of virus mutations in only one transmission step. For example, the samples named OY and AU represent a married couple (IT2; Fig. 1) who were probably infected by the same individual, as they presented with clinical symptoms about the same time. Not surprisingly, we did not identify any sequence differences in the viruses from these two individuals. Similarly, we did not detect any sequence variations between viruses in samples collected 10 days apart from the same individual (IT1, I7 and 7R, Fig. 1). These results not only confirm the rather slow mutation rate of SARS-CoV-2 but also demonstrate the limitations of using molecular epidemiology applications in tracing virus transmission.

Mutations in the Nsp5A ORF, encoding the main proteinase Mpro

In addition to the mutations described above, we also identified a unique (as of September 30, 2020, GISAID.org), high-confidence (3866x) mutation (V13I) within the Nsp5A ORF, encoding the main proteinase Mpro, in one of

the samples (8U, Supplementary Table 2 and Fig. 2). It is important to mention that the patient 8U had not received any antiviral treatment. Remarkably, we did not detect the same mutation in samples from epidemiologically linked patients who were likely to have transmitted the virus to 8U (not shown), suggesting that a *de novo* mutation might have occurred during that transmission. The analysis of structural models of the Nsp5A V13I mutation showed low mobility for the V13 position, suggesting that the mutation would have a negligible impact on protein function (e.g., conformational changes that drive catalysis) or selective advantage for the virus. Nonetheless, it is notable that mutations within the protease gene might influence its sensitivity to protease inhibitors, as has been observed with HIV and lopinavir/ritonavir [29, 30]. Thus, it would be very interesting to investigate if this particular mutation has spread further within the population. Indeed, during the revision of this manuscript, 13 additional viruses from four different countries (0.01% of all samples with Nsp5 sequences; GISAID.org) were identified with the same mutation.

In conclusion, we developed an in-house SARS-CoV-2 sequencing panel and obtained the first sequences of SARS-CoV-2 circulating in Croatia during early spring of 2020. This work was important to present national capacities and readiness to contribute to the global effort in understanding the COVID-19 pandemic [12]. Although the scope was rather limited (21 samples), by merging direct sequencing with an epidemiological approach, we found that patients presenting with symptoms who had returned from other countries, including Turkey, Italy, Austria, and the Netherlands, were infected with diverse virus variants belonging to different clades. As expected, these results confirm that widespread transmission had already occurred in March 2020 [1, 3, 23, 25]. The 2020 spring wave of the pandemic was, to an extent, contained by a myriad of epidemiological control measures, but clearly, the pandemic will not be resolved until there is a broad distribution of vaccines. However, at the same time, the virus continues to evolve, and there is great uncertainty about how newly arising mutations will affect the course of the pandemic. At this point, it is important to continuously monitor the evolution of the virus, because new, possibly dangerous variants might emerge. Also, it is equally important to complement the sequence information with metadata (e.g., symptoms, severity, treatment, outcome, and transmission) to be able to assess the functional importance of a particular mutation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-021-05029-7.

References

1. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole A, Pereyaslov D, WHO European Region sequencing laboratories and GISAID EpiCoV group (2020) Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill 25:2001410
2. Anson B, Ghosh AK, Mesecar A (2020) X-ray Structure of SARS-CoV-2 main protease bound to GRL-024-20 at 1.45 A. https://www.rcsborg/structure/6xr3
3. Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, Mechain M, Meurice L, Nguyen M, Bassi C, Yamani E, Behillil S, Ismael S, Nguyen D, Malvy D, Lescure FX, Georges S, Lazarus C, Tabai A, Stempfelet M, Enouf V, Coignard B, Levy-Bruhl D, Investigation Team (2020) First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. Euro Surveill 25:2000094
4. Bohmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, Bohm S, Woudenberg T, Ackermann N, Konrad R, Eberle U, Treis B, Dangel A, Bengs K, Fingerle V, Berger A, Hormansdorfer S, Ippisch S, Wicklein B, Grahl A, Portner K, Muller N, Zeitlmann N, Boender TS, Cai W, Reich A, An der Heiden M, Rexroth U, Hamouda O, Schneider J, Veith T, Muhlemann B, Wolfel R, Antwerpen M, Walter M, Protzer U, Liebl B, Haas W, Sing A, Drosten C, Zapf A (2020) Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. Lancet Infect Dis 20:920–928
5. Day T, Gandon S, Lion S, Otto SP (2020) On the evolutionary epidemiology of SARS-CoV-2. Curr Biol 30:R849–R857
6. Elbe S, Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 1:33–46
7. Giovanetti M, Angeletti S, Benvenuto D, Ciccozzi M (2020) A doubt of multiple introduction of SARS-CoV-2 in Italy: A preliminary overview. J Med Virol 92:1634–1636
8. Gomez-Carballa A, Bello X, Pardo-Seco J, Martinon-Torres F, Salas A (2020) Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res 30:1434–1448
9. Hossain MP, Junus A, Zhu X, Jia P, Wen TH, Pfeiffer D, Yuan HY (2020) The effects of border control and quarantine measures on the spread of COVID-19. Epidemics 32:100397
10. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
11. Islam MR, Hoque MN, Rahman MS, Alam A, Akther M, Puspo JA, Akter S, Sultana M, Crandall KA, Hossain MA (2020) Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Rep 10:14004
12. Jurak I, Rukavina T, Vugrek O (2020) Successful sequencing of the first SARS-CoV-2 genomes from Croatian patients. Croat Med J 61:302–303
13. Kaushal N, Gupta Y, Goyal M, Khaiboullina SF, Baranwal M, Verma SC (2020) Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. Pathogens 9:565
14. Kinross P, Suetens C, Gomes Dias J, Alexakis L, Wijermans A, Colzani E, Monnet DL, ECDC Public Health Emergency Team (2020) Rapidly increasing cumulative incidence of coronavirus disease (COVID-19) in the European Union/European Economic Area and the United Kingdom, 1 January to 15 March 2020. Euro Surveill 25:2000285
15. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley

B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Sheffield C-GG, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC (2020) Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182(812–827):e819
16. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874
17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
18. Lorenzo-Redondo R, Nam HH, Roberts SC, Simons LM, Jennings LJ, Qi C, Achenbach CJ, Hauser AR, Ison MG, Hultquist JF, Ozer EA (2020) A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. EBioMedicine 62:103112
19. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641
20. Mercatelli D, Giorgi FM (2020) Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol 11:1800
21. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY (2020) Spike mutation D614G alters SARS-CoV-2 fitness. Nature. https://doi.org/10.1038/s41586-020-2895-3
22. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407
23. Rito T, Richards MB, Pala M, Correia-Neves M, Soares PA (2020) Phylogeography of 27,000 SARS-CoV-2 genomes: Europe as the major source of the COVID-19 pandemic. Microorganisms 8:1678
24. Sjodin H, Wilder-Smith A, Osman S, Farooq Z, Rocklov J (2020) Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Euro Surveill 25:2000280
25. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, Fabiani C, Anselmo A, Ciammaruconi A, Fortunato A, De Santis R, Fillo S, Capobianchi MR, Gismondo MR, Ciervo A, Rezza G, Castrucci MR, Lista F, On Behalf of Iss Covid-Study G (2020) Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. Euro Surveill 25:2000305
26. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT, Balloux F (2020) Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol 83:104351
27. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol 92:667–674
28. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269
29. Xue X, Yu H, Yang H, Xue F, Wu Z, Shen W, Li J, Zhou Z, Ding Y, Zhao Q, Zhang XC, Liao M, Bartlam M, Rao Z (2008) Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J Virol 82:2515–2527
30. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368:409–412
31. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, Boerwinkle E, Fu YX (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. BMC Evol Biol 4:21

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.