*Briefings in Bioinformatics*, 00(00), 2021, 1–3

**doi: 10.1093/bib/bbab129**
Advance Access Publication Date: 00 Month 0000
Letter to Editor

# Letter to the Editor: Significant mutation enrichment in inverted repeat sites of new SARS-CoV-2 strains

Martin Bartas, Pratik Goswami, Matej Lexa, Jiří Červeň, Adriana Volná, Miroslav Fojta, Václav Brázda and Petr Pečinka

Corresponding authors. Václav Brázda, Department of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic E-mail: vaclav@ibp.cz; Petr Pečinka, Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic E-mail: petr.pecinka@osu.cz

## Abstract

In a recently published paper, we have found that SARS-CoV-2 hot-spot mutations are significantly associated with inverted repeat loci and CG dinucleotides. However, fast-spreading strains with new mutations (so-called mink farm mutations, England mutations and Japan mutations) have been recently described. We used the new datasets to check the positioning of mutation sites in genomes of the new SARS-CoV-2 strains. Using an open-access Palindrome analyzer tool, we found mutations in these new strains to be significantly enriched in inverted repeat loci.

**Key words:** SARS-CoV-2; mutations; inverted repeats

SARS-CoV-2, a member of *Coronaviridae* family, has caused recent COVID-19 pandemic. In our previously published research [1], we have found that SARS-CoV-2 hot-spot mutations are significantly associated with inverted repeat loci (IRs) and CG dinucleotides. In the meantime, three novel alarming datasets of SARS-CoV-2 hot-spot mutations have been found: so-called mink farm dataset of recurrent SARS-CoV-2 mutations (16 November 2020) [2]; nonsynonymous mutations and deletions inferred to occur on the branch leading to B.1.1.7 'England' lineage (19 December 2021) [3]; and lineage-defining mutations

**Martin Bartas** is a postdoc in the Department of Biology, University of Ostrava, Czech Republic. His research interests include noncanonical forms of nucleic acids, protein interactions and bioinformatics.
**Pratik Goswami** is carrying out his PhD work at the Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic. He is a student of the Faculty of Science, Masaryk University, Brno, Czech Republic. His research focus includes study of nucleic acids structures and their interactions with proteins.
**Matej Lexa** is a Researcher at the Faculty of Informatics, Masaryk University, Brno, Czech Republic. His research focus includes plant biology, bioinformatics and transposable elements.
**Jiří Červeň** works as a research fellow in the Department of Biology, University of Ostrava, Czech Republic. His work spans molecular biology and microbiology.
**Adriana Volná** is a PhD student in the Department of Physics, University of Ostrava, Czech Republic. Her work spans molecular virology, plant biology and interdisciplinary approaches.
**Miroslav Fojta** is an Assistant Professor and the Head of the Department of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic. He is involved in studies of DNA structures in solution and at surfaces, DNA damage and chemical modification, and DNA-protein interactions.
**Václav Brázda** is an Assistant Professor and the Head of the Laboratory of Protein–DNA Interactions, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic. He is studying the interaction of proteins with DNA, local DNA structures and p53 protein and is a co-author of a web-bioinformatics server (bioinformatics.ibp.cz).
**Petr Pečinka** is an Assistant Professor and the Team Leader of the Molecular Biology group in the Department of Biology, University of Ostrava, Czech Republic. The University of Ostrava is a public research university educating nearly 9000 students in six faculties and provides an international environment in which to study. The university is proud to retain lecturers, professors and researchers that are leading figures in their fields of expertise whom are scientists and inspirational and open-minded personalities with a vivid sense of creativity.
**Submitted:** 8 February 2021; **Received (in revised form):** 16 March 2021

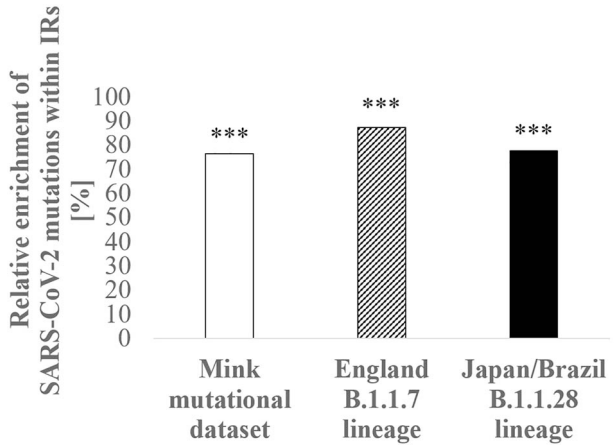


**Figure 1**. Relative percentual enrichment of SARS-CoV-2 mutational datasets within IRs. The assumption is that there is no enrichment of IRs within SARS-CoV-2 hotspots (zero value in the plot). A one-sample *t*-test was used to statistically compare the number of real SARS-CoV-2 mutations within IRs with random mutations localization (done in 100 replicates). *** indicates *P*-value < 0.001. Detailed information can be found in [1].

of the different B.1.1.28 clades detected in Brazil and Japan (11 January 2021) [4]. We have analyzed these new datasets of SARS-CoV-2 mutations by the Palindrome analyzer [5] and found a remarkably high and statistically significant (*P*-value < 2.2e−16) enrichment of mutations within the IRs in all three new SARS-CoV-2 strains (Figure 1). If we consider only the longer IRs (7+ nucleotides), the observed enrichment is even two-times higher (for all analyzed mutational datasets and overlays of IRs with tested mutations, see Supplementary Data available online at https://academic.oup.com/bib).

It is therefore evident that inverted repeat loci play an important role in SARS-CoV-2 genetic drift and should be extra monitored by predictive analyses and modelling. Our dataset is also in line with previously published data [6] saying that SARS-CoV-2 mutations are strongly biased toward C > U and U > C transitions (32 out of 67 analyzed mutations; 47.8% of all SARS-CoV-2 mutations in the novel datasets belong to C > U + U > C transitions, while the expected value is slightly above 15% for C > U + U > C in the SARS-CoV-2 genome [6]).

**Key Points**

- We have analyzed three novel datasets of SARS-CoV-2 hot-spot mutations.
- IRs are a rich source of SARS-CoV-2 genetic instability.
- These novel results can be further utilized for predictive analyses of further possible SARS-CoV-2 mutations.

## Supplementary data

Supplementary data are available online at https://academic.oup.com/bib.

## Funding

The Czech Science Foundation (18-15548S, 18-18699S); and the SYMBIT project Reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed from the ERDF.

## Data availability

All data are available in the paper and in the Supplementary data.

## References

1. Goswami P, Bartas M, Lexa M, *et al*. SARS-CoV-2 hot-spot mutations are significantly enriched within inverted repeats and CpG island loci. *Brief Bioinform* 2020;**22**:1338–1345.
2. van Dorp L, Tan CC, Lam SD, *et al*. Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv* 2020. doi: https://doi.org/10.1101/2020.11.16.384743.
3. Rambaut A, Loman N, Pybus O, *et al*. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*. virological.org (https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563) 2020.
4. Naveca F, Nascimento V, Souza V, *et al*. *Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein*. virological.org (https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585) 2021.
5. Brázda V, Kolomazník J, Lýsek J, *et al*. Palindrome analyser—a new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. *Biochem Biophys Res Commun* 2016;**478**:1739–45.
6. Matyášek R, Kovařík A. Mutation patterns of human SARS-CoV-2 and bat RaTG13 coronavirus genomes are strongly biased towards C > U transitions, indicating rapid evolution in their hosts. *Genes* 2020;**11**:761.