In silico detection of (CRISPR) spacers matching *Betacoronaviridae* genomes in gut metagenomics sequencing data

Leoni G.[1,2], Petrillo M.[2], Puertas-Gallardo A.[2], Sanges R.[1], Patak A.[2]

1. Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste (Italy);
2. Joint Research Center (JRC), Ispra (Italy).

The CRISPR-Cas system is the major component of the prokaryotic adaptive immune system (Horvath & Barrangou, 2010). CRISPR, which stand for "Clustered Regularly Interspaced Short Palindromic Repeats", are genomics arrays found in the DNA of many bacteria. They consist in short repeated sequences (size 23-47 base pairs), separated by unique sequences of similar length (spacers), that often derives from phages and viral infections, plasmids or mobile genetic elements (Shmakov et al., 2017). CRISPRs are coupled to specific "CRISPR-associated genes" (Cas) to form the so called CRISPR-Cas system. This system has the primary role to protect prokaryotes from virus and other mobile genetic elements activity by conferring immunological memory from past infections (Garneau et al., 2010; Nussenzweig & Marraffini, 2020).

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) is a single-stranded RNA virus that rapidly emerged in 2019. In humans, it causes coronavirus disease 2019 (COVID-19), an influenza-like disease that is primarily thought to infect the lungs with transmission through the respiratory route. However, clinical evidence suggests that the intestine may present another viral target organ, a potential hiding place for the virus, which may explain the persistence of COVID-19 symptoms after months from patients recovery (Lamers et al., 2020). Furthermore, extra-pulmonary clinical manifestations of COVID-19 are reported. Nonetheless, although a link between SARS-CoV-2 infection and the mis-regulation of the gut microbiome was suggested, its involvement remains largely unexplored (Brooks & Bhatt, 2021).

To simultaneously verify both the potential existence of SARS-CoV-2 in gut and to test whether the human gut microbiome may be stressed by SARS-CoV-2 infection, we developed a bioinformatic workflow based on the detection of *Betacoronaviridae-specific* CRISPR spacers from ~28,000 public available gut metagenomics data. To process such "Big Biological Data" in a reasonable CPU time, we relied on a HTCondor High Throughput Computing System, characterized by 10 Tflops of computing capacity and more than 80 Tbytes of storage. Computing block was composed by 8 Nodes IBM x3550 with two Intel Xeon processor E5-2600 v3 product family CPUs with10 cores 2.6 GHz, two QPI links up to 9.6 GT/s each and 256 GB of RAM. While our work is still ongoing, preliminary results revealed the presence of some *Betacoronavirus-specific* spacers in the human gut metagenomics data, proving that SARS-like viruses can target human gut and suggesting that the human microbiome can be stressed by the systemic viral infection. By collecting further data, we aim to strengthen our results as well as to investigate the effects of the SARS-COV-2-induced microbiome stress to the host.

Bibliography

Brooks, E. F., & Bhatt, A. S. (2021). The gut microbiome: A missing link in understanding the gastrointestinal manifestations of COVID-19? *Molecular Case Studies*, *7*(2), a006031. https://doi.org/10.1101/mcs.a006031

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*. https://www.science.org/doi/abs/10.1126/science.1179555

Lamers, M. M., Beumer, J., Vaart, J. van der, Knoops, K., Puschhof, J., Breugem, T. I., Ravelli, R. B. G., Schayck, J. P. van, Mykytyn, A. Z., Duimel, H. Q., Donselaar, E. van, Riesebosch, S., Kuijpers, H. J. H., Schipper, D., Wetering, W. J. van de, Graaf, M. de, Koopmans, M., Cuppen, E., Peters, P. J., … Clevers, H. (2020). SARS-CoV-2 productively infects human gut enterocytes. *Science*. https://www.science.org/doi/abs/10.1126/science.abc1669

Nussenzweig, P. M., & Marraffini, L. A. (2020). Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annual Review of Genetics*, *54*(1), 93–120. https://doi.org/10.1146/annurev-genet-022120-112523

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *mBio*, *8*(5), e01397-17. https://doi.org/10.1128/mBio.01397-17