## Supplement of the paper

# “The emergence of high-fitness variants accelerates the slowdown of genome heterogeneity in the coronavirus”

Since the outbreak of the COVID-19 pandemic, the SARS-CoV-2 coronavirus accumulated an important amount of genome variability through mutation and recombination. To test evolutionary trends that could inform us on the adaptive process of the virus to its human host, we compute a genome-wide measure of Sequence Compositional Complexity (*SCC*) in high-quality coronavirus genomes from across the globe, covering the full span of the pandemic. In early samples, we find no statistical support for any trend in *SCC* values over time, although the virus genome appears to evolve faster than Brownian Motion expectation. However, in samples taken after the emergence of Variants of Concern with higher transmissibility, and controlling for phylogenetic and sampling effects, we detect a declining trend for *SCC* and an increasing one for its absolute evolutionary rate. This means that the decline in *SCC* itself accelerated over time, and that increasing fitness of variant genomes lead to a reduction of their genome sequence heterogeneity.

# Supplementary files

|  |  |
| --- | --- |
| **File** | **Description** |
| SupplementaryTables S1-S18.xlsx | The strain name, the collection date, and the SCC values for each analyzed genome. |
| SupplementaryTableS19.pdf | A complete list acknowledging all originating and submitting laboratories for the sequence data in GISAID EpiCoV on which these analyses are based. |
| SupplementaryTable S20.pdf | A complete list acknowledging the authors, originating and submitting laboratories of the genetic sequences we used for the analysis of the Nextstrain sample. |
| PhylogeneticTimetrees\_NexusFormat.zip | Phylogenetic timetrees (Nexus format). |
| PhylogeneticTimetrees\_NewickFormat.zip | Phylogenetic timetrees (Newick format). |
| SCCdata.zip | SCC data. |