Multi-organ spread and intra-host diversity of SARS-CoV-2 support viral persistence, adaptation, and a mechanism that increases evolvability

Julieta M. Manrique[1,2], Santiago Maffia-Bizzozero[3], M. Victoria Delpino[1,4], Jorge Quarleri[1,4] & Leandro R. Jones[1,2,*]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avenida Rivadavia 1917, C1083ACA Ciudad Autónoma de Buenos Aires, Argentina.
[2] Laboratorio de Virología y Genética Molecular (LVGM), Facultad de Ciencias Naturales y Ciencias de la Salud, Universidad Nacional de la Patagonia San Juan Bosco, Belgrano 160, (9100) Trelew, Argentina .
[3] Morgue Judicial de la Nación, Ciudad Autónoma de Buenos Aires, Argentina.
[4] Laboratorio de Inmunopatología Viral, Instituto de Investigaciones Biomédicas en Retrovirus y Sida (INBIRS), Facultad de Ciencias Médicas, Universidad de Buenos Aires, Argentina.

* Corresponding author: Email: lrj000@gmail.com; ljones@conicet.gov.ar.

Supplementary Figure Legends

Figure S1. Bioinformatic workflow. Demultiplexed reads (*bcl-convert* tool) were quality controlled using *fastp* (*QC reads*) and aligned against MN908947.3 reference sequence by Burrows-Wheeler alignment in *bwa* (*BW alignment*), for both SNV calling with regard to the reference strain using *ivar* (*SNV calling*), and generating consensus sequences (*Organ Consensus)* for each specimen studied (Table S1) with *samtools*. Organ consensus sequences from each autopsy were aligned with each other using *MAFFT*, and the aligned organ genomes were used to obtain consensus sequences for each studied corpse by *em_cons* (*Case Consensus*). In order to identify intra-host polymorphisms, cleaned reads were re-aligned to case consensus sequences, and outputs were used for identifying intra-host mutations by SNV (relative to infecting lineage) calling. Workflows were implemented in *bash* scripts. Single nucleotide variants (SNV) identified by *ivar* were filtered using an *ad-hoc* R script (*SNV Filters*). Final products were analyzed directly, *i.e.* using spreadsheet software and/or base functions of the R statistical programming language, or processed by the high-level applications *UpSetR* and *Circos*.

Figure S2. Polymorphisms retrieved from specimens 13-16 of case 5. The figure is organized into quadrants, each of which corresponds to a different organ specimen as indicated in the outermost part of the graphic (further details are provided in Table S1). First and second tracks display the viral genes and the corresponding gene products. Locations in the genome are indicated on the first track, where each tick indicates a 1Kb unit. The innermost tracks consist of segments (gray) representing the whole viral genome. The location if intra-host polymorphisms are indicated by points on the segments, using colors to indicate whether they correspond to synonymous (blue) or non-synonymous (red) mutations. Grids on the segments are used to display alternative allele frequencies. Each grid line correspond to 0.2 (20%) abundance units. To facilitate locating the polymorphisms in relation to the viral genes and proteins, a genome scale is provided below each segments, with ticks separated from each other by 1Kb as in the outermost scale described above. The links (lines) in the center of the figure unite polymorphic positions of the genome affected by same iSNVs. Colored links correspond to polymorphisms detected in all cases (details in Figure 1).

Figure S3. Polymorphisms retrieved from specimens 42-46 of case 18. Data are presented as in Figure S2.

Figure S4. Polymorphisms retrieved from specimens 52-56 of case 24. Data are presented as in Figure S2.

Figure S5. Polymorphisms retrieved from specimens 62-66 of case 28. Data are presented as in Figure S2.