SUPPLEMENTARY MATERIAL

Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene

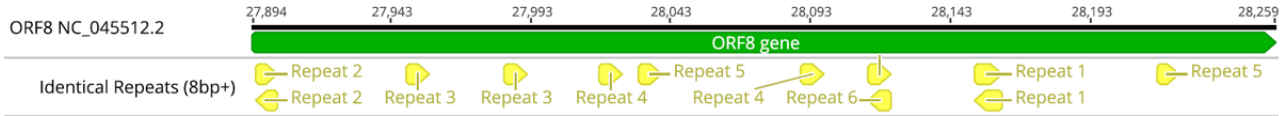
Filipe Pereira

Departamento de Ciências da Vida - Universidade de Coimbra. Calçada Martim de Freitas, 3000-456 Coimbra, Portugal.

IDENTIFICA, Science and Technology Park of the University of Porto - UPTEC, Rua Alfredo Allen, N.º455/461. 4200-135 Porto, Portugal.

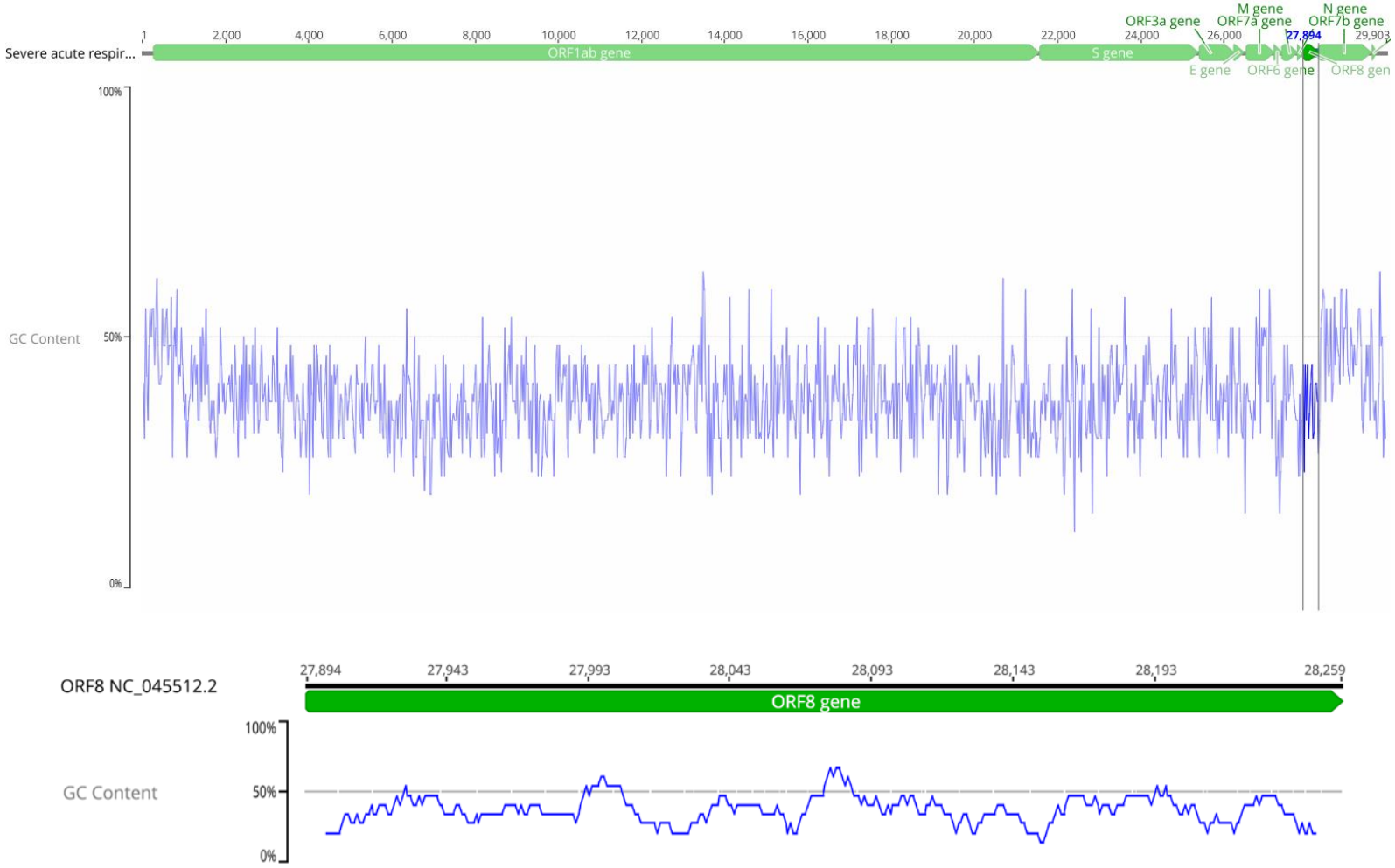
Corresponding author: E-mail: fpereirapt@gmail.com

Supplementary Figure S1. Perfect repeats with more than 8 nucleotides identified in the SARS-CoV-2 reference genome (NC_045512.2).

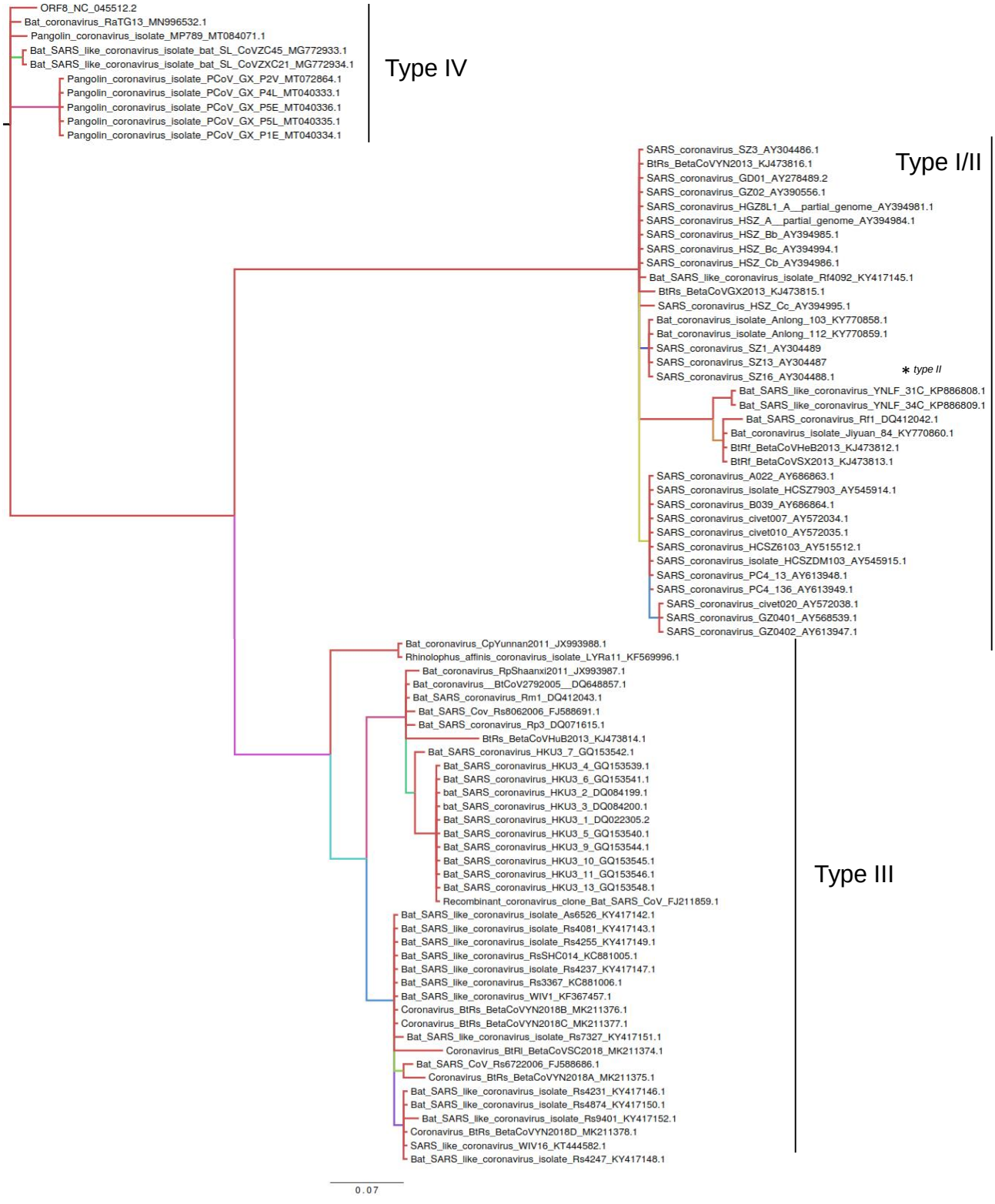


Repeat number	Minimum	Maximum	Length	Direction
1	260	269	10	forward
1	260	269	10	reverse
2	3	10	8	forward
2	3	10	8	reverse
3	92	99	8	forward
3	57	64	8	forward
4	198	205	8	forward
4	126	133	8	forward
5	325	332	8	forward
5	140	147	8	forward
6	222	229	8	forward
6	222	229	8	reverse

Supplementary Figure S2. GC content across the SARS-CoV-2 reference genome (NC_045512.2) and ORF8 locus calculated with a sliding window of 15 nucleotides.



Supplementary Figure S3. *Betacoronaviruses* lineage B phylogeny based on ORF8 protein sequences. The Bayesian phylogenetic tree was built with the reference SARS-CoV-2 and all bat coronaviruses, pangolin coronaviruses and SARS-related sequences identified in blast searches (*n*=84). Bayesian posterior probabilities are shown on basal nodes and as colours in branches (from high values in red to low values in violet). The scale bar indicates substitutions per site.



Supplementary Figure S4. Multiple sequence alignment of ORF8 proteins in betacoronaviruses lineage B. The consensus sequence and identity is indicated for every position above the alignment, with high and low identity values represented by green and red bars, respectively.