



Diversidade genética do novo coronavírus SARS-CoV-2 (COVID-19) em Portugal

Mais informações em <https://insaflu.insa.pt/covid19/>

Relatório de situação

12 de Maio de 2020

- O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **436 sequências do genoma do novo coronavírus SARS-CoV-2**, obtidas de amostras colhidas em 44 laboratórios/hospitais representando 77 concelhos (**Figura 1**).
- A maioria dos vírus (87.8 %) pertence ao “clade” **A2, subclade A2a**, o mais representado em toda a Europa (**Figura 2**). O “clade” **A2 é caracterizado pela mutação “aa D614G” (“nt A23403G”) na proteína “Spike (S)” (Figura 3)**. Esta proteína é responsável pela entrada do vírus SARS-CoV-2 nas células humanas, sendo também o principal antígeno deste vírus pandémico. O potencial impacto da mutação D614G na capacidade de transmissão / virulência do SARS-CoV-2 está a ser alvo de monitorização detalhada pela comunidade científica.
- Outros clades também detectados incluem: **A1a (5.9 %), A6 (0.5 %), B (3.7 %), outros perfis (2.1 %) (Figura 2)**. Destaca-se que maioria dos genomas do **clade B (8/16)** foi detectada em **Évora**. Este clade apresenta uma maior frequência em Espanha, apontando que a sua introdução no Alentejo poderá ter tido origem neste país. Por sua vez, o **clade A6**, frequente nos Países Baixos, foi detectado exclusivamente na **Ilha da Madeira**, sugerindo uma introdução diferente dos restantes casos observados no arquipélago (<https://nextstrain.org/ncov/europe>).
- Apesar da grande variabilidade de perfis mutacionais, observam-se já **clusters filogenéticos congruentes com dados epidemiológicos (ex, geografia, contactos confirmados, etc) e outros sugestivos de potencial link epidemiológico**. Destaca-se um grande cluster constituído pela maioria dos genomas avaliados do concelho de Ovar (**Figura 4**). Este resultado lança a hipótese de que a maioria dos casos de COVID-19 nesta região tenha resultado de uma única introdução do vírus (início de Março), desencadeando uma grande cadeia de transmissão neste concelho. Resultados preliminares apontam que este foco poderá estar relacionado com o foco inicial no Concelho de Felgueiras e ter progredido para outros locais, em particular o Distrito de Viseu, onde se detectaram já alguns genomas do mesmo sub-clade genético.
- O **número médio de mutações por genoma** (comparando com o primeiro genoma sequenciado na China; MN908947.3) **é de 8 (variando entre 2 e 13) mutações (Figura 5)**, o que se enquadra dentro da taxa de mutação prevista para este vírus (i.e., cerca de 2 mutações por genoma por mês).
- As mutações distribuem-se ao longo dos 29903 nucleótidos que constituem o genoma do novo coronavírus SARS-CoV-2 (**Figura 6**), tendo sido já detectadas mais de 350 mutações distintas no conjunto dos 436 genomas analisados até à data.
- 31 mutações não-sinónimas alteram a proteína “Spike (S)” (**Figura 3**), sendo que apenas duas delas (**D614G e D839Y**) estão presentes em mais de 20% dos vírus analisados (87.8% e 22%, respectivamente). Curiosamente, enquanto que a mutação **D614G caracteriza o clade A2 altamente representado na Europa (Figura 3)**, a mutação **D839Y é exclusiva do cluster relativo à cadeia de transmissão mais evidente até agora, com foco em Ovar (Figura 4)**.

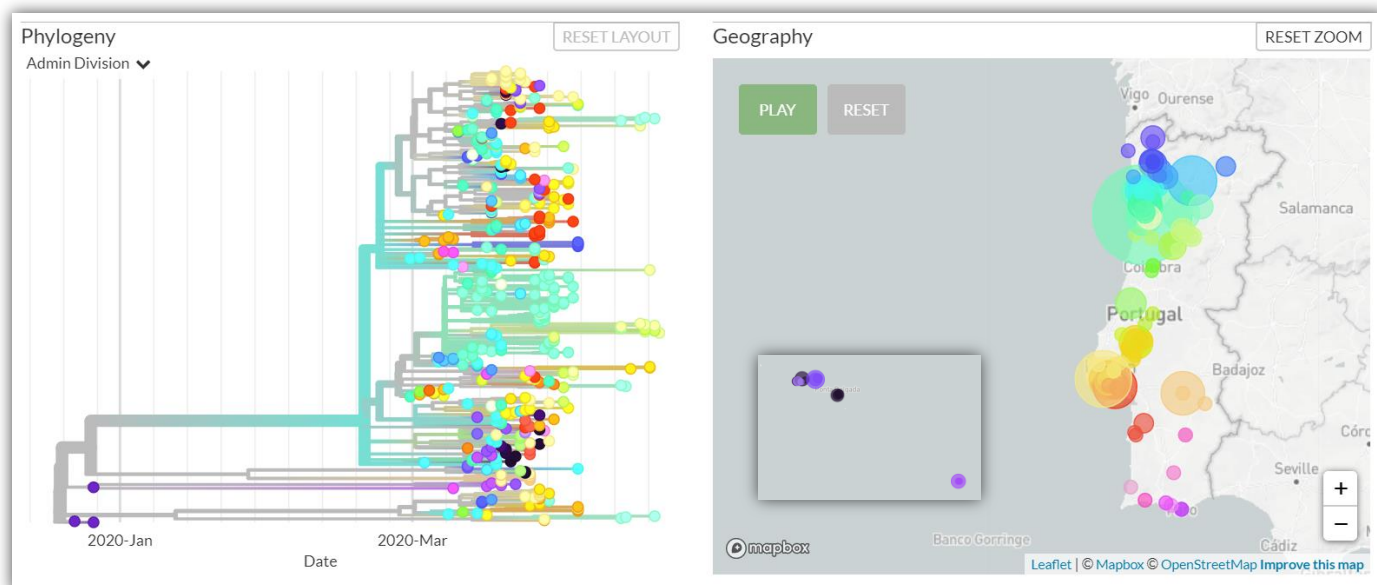


Figura 1. Visão global da diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal. Os diferentes genomas (representados por círculos no painel à esquerda) estão coloridos de acordo com o local de residência, com a mesma tonalidade no mapa – o tamanho dos círculos no mapa é proporcional ao número de genomas sequenciados por localidade (consultar o site <https://insaflu.insa.pt/covid19/> para mais detalhes).

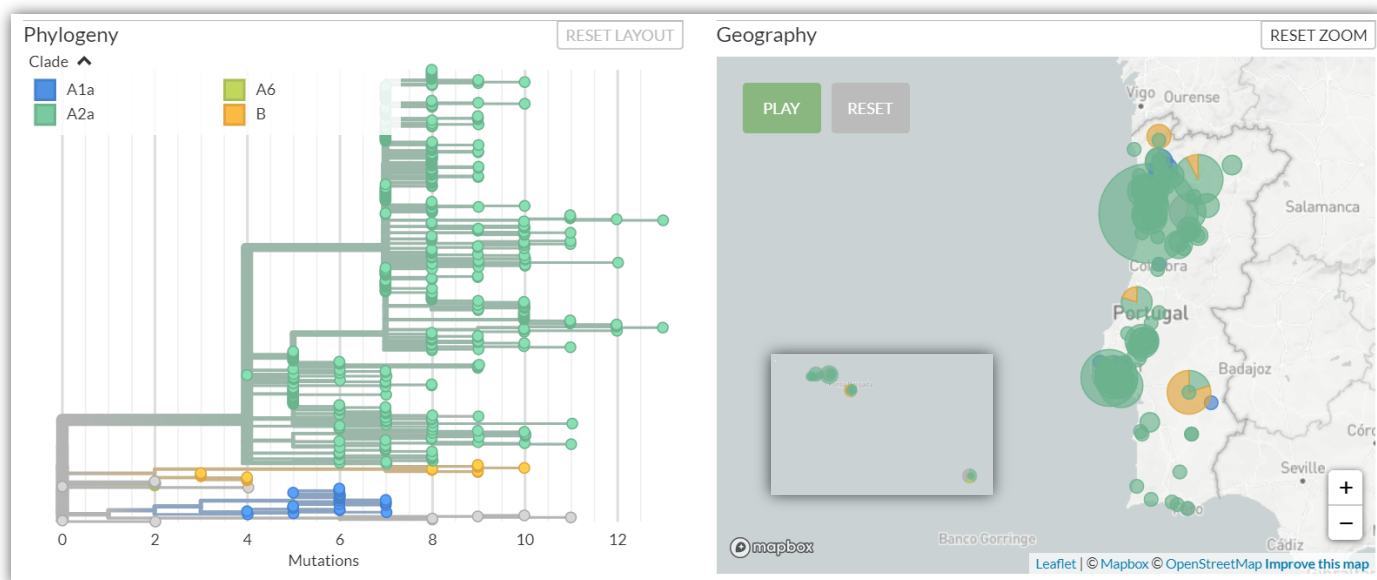


Figura 2. Diversidade genética e dispersão geográfica do vírus SARS-CoV-2 por “clade” genético (tal como definido em <https://nextstrain.org/>).

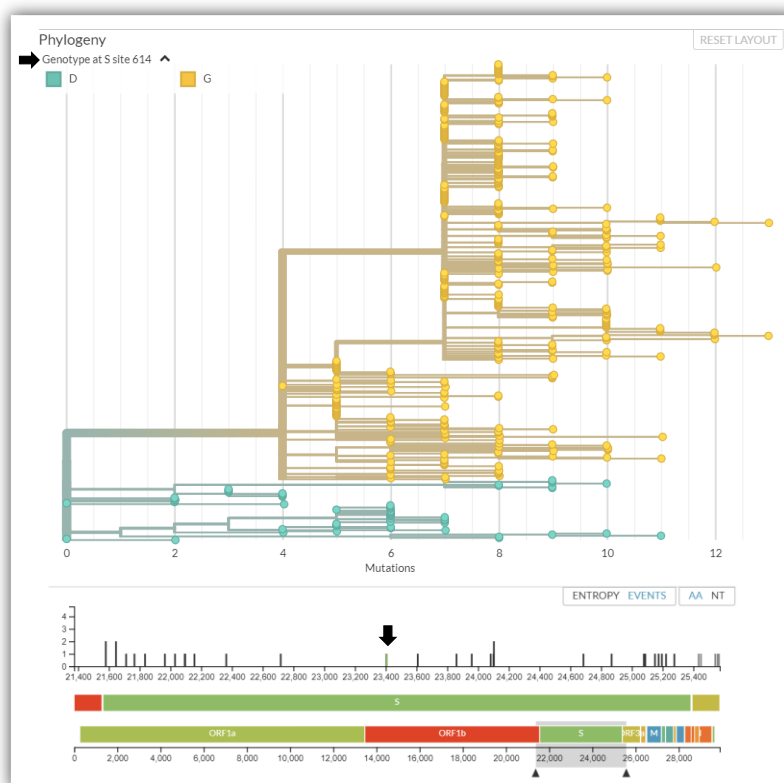


Figura 3. Árvore filogenética colorida de acordo com os variantes (D ou G) na posição 614 (correspondente à alteração nucleotídica A23403G) da proteína “Spike (S)”. Os vírus que apresentam a variante 614G pertencem ao “clade” mais representado em toda a Europa. Abaixo da árvore estão sinalizadas as posições para as quais já se observaram alterações aminoacídicas na proteína “Spike (S)” nos vírus a circular em Portugal.

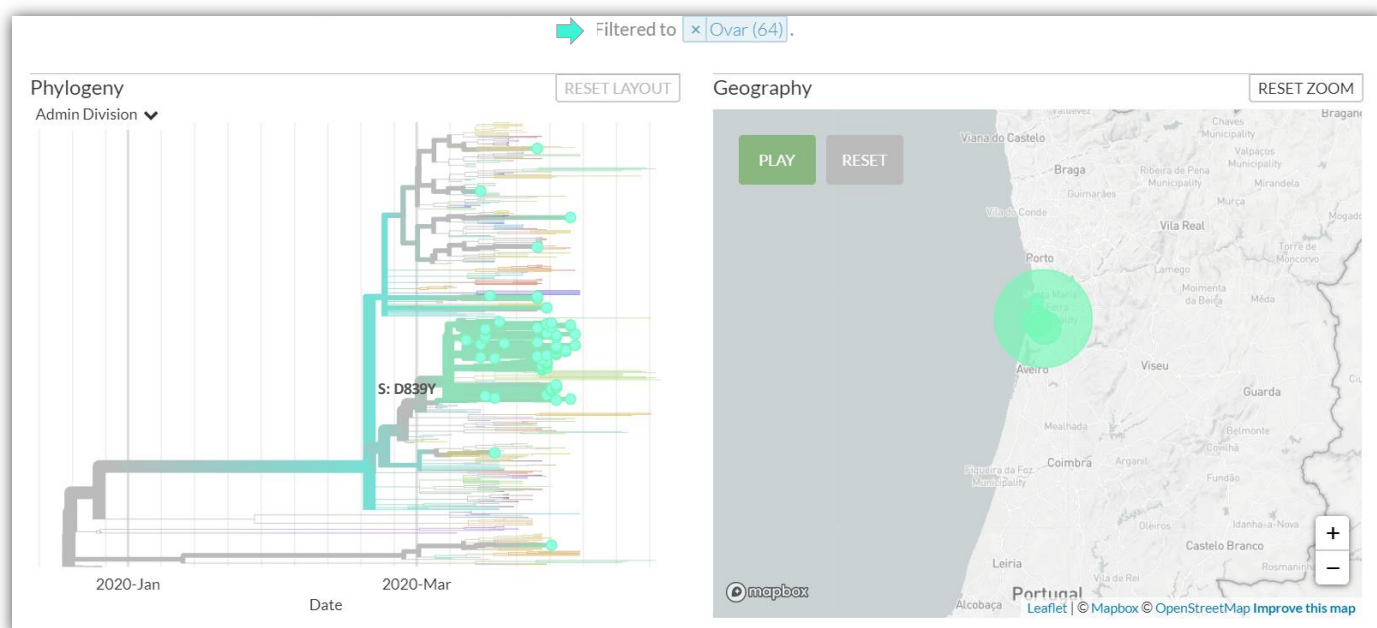


Figura 4. Diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal, destacando os genomas avaliados do concelho de Ovar. A grande maioria agrupa num único sub-clade filogenético o qual tem como “marcador” a mutação D839Y na proteína “Spike (S)”.

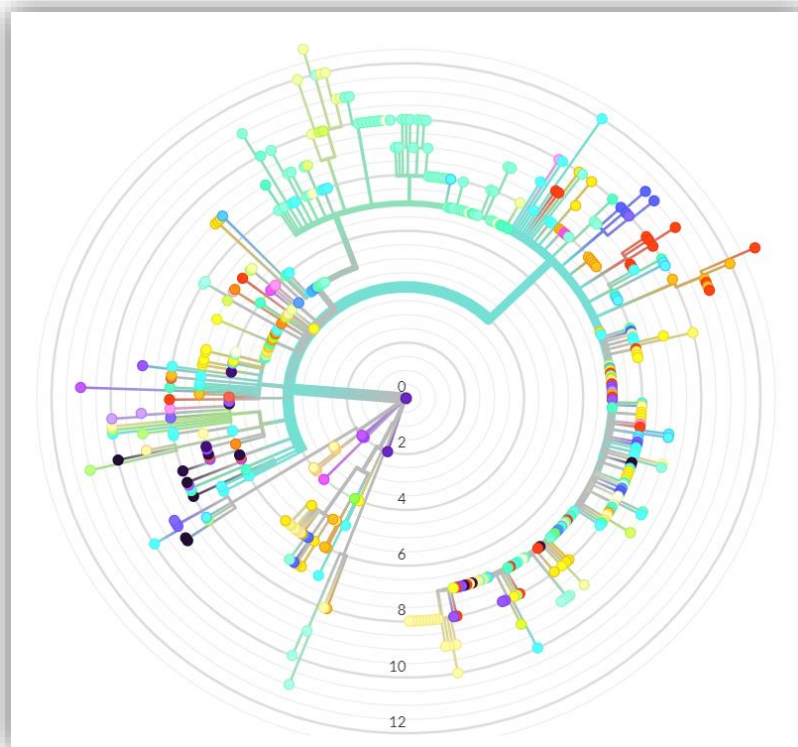


Figura 5. Árvore filogenética radial em que a distância de cada genoma ao centro é proporcional ao número de mutações observadas em comparação com o primeiro genoma de SARS-CoV-2 reportado na China (os diferentes genomas estão representados por círculos e coloridos coloridos de acordo com o concelho de residência).

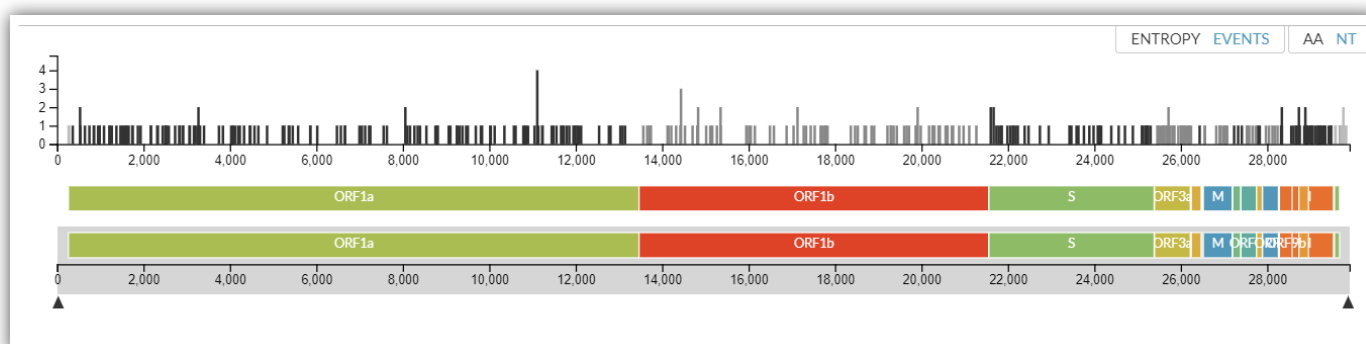


Figura 6. Representação gráfica do genoma do novo coronavírus SARS-CoV-2 (com diferentes cores para diferentes genes), em que, no painel acima, estão sinalizadas as posições para as quais já se observaram alterações nucleotídicas (mutações) nos vírus a circular em Portugal.



NOTAS ADICIONAIS (também disponível em <https://insaflu.insa.pt/covid19/>)

Objectivos gerais do estudo:

- Determinação dos **padrões de disseminação do vírus nas diferentes regiões de Portugal** e em diferentes grupos populacionais.
- Determinação dos perfis mutacionais do SARS-CoV-2 para **identificação e monitorização de cadeias de transmissão**, bem como **identificação de novas introduções do vírus** em Portugal.
- Prever o início da transmissão na comunidade e aferir o impacto das medidas de contenção, avaliando a **contribuição da transmissão local versus importações do vírus**.
- Determinação do **grau de variabilidade genética de antígenos ou alvos de fármacos antivirais** com possível impacto no desenvolvimento / eficiência de medidas profiláticas (vacinas) e terapêuticas.
- Determinação de possíveis **associações entre perfis genéticos (mutacionais)** do SARS-CoV-2 e **determinadas manifestações clínicas** (ex. diferentes graus de severidade da COVID-19) ou diferente **capacidade de transmissão** do vírus.
- Estudar os **mecanismos evolutivos do vírus** e a sua relação com os perfis de disseminação em diferentes regiões de Portugal e em diferentes grupos populacionais.
- **Contribuir para a avaliação da relevância funcional e fenotípica** de mutações particulares.

Métodos

- **Procedimento Pre-NGS:** adaptado da Artic Network (<https://artic.network/ncov-2019>, <https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w>)
- **Procedimento NGS:** Nextera XT e MiSeq (Illumina)
- **Dos "reads" às sequências do genoma:** [INSaFLU](#)
- **Das sequências do genoma à "filogeografia*":** Nextstrain (mais detalhes sobre o método podem ser encontrados em <https://nextstrain.org/ncov> e aqui <https://github.com/nextstrain/ncov>)

* O posicionamento geográfico reflecte o local (Concelho – "Admin Division" ou Freguesia – "Location") de residência ou, caso não exista informação, local de ocorrência ou da entidade que enviou a amostra. Apenas são indicadas as freguesias com população residente superior a 5000 pessoas (Fonte: CENSOS 2011 - Instituto Nacional de Estatística). Para as restantes freguesias, por motivos de confidencialidade, é apenas indicado o Concelho.

Agradecimentos

- A todos os laboratórios nacionais que enviam amostras clínicas (suspeitas ou positivas para SARS-CoV-2) para o Laboratório Nacional de Referência da Gripe e outros vírus respiratórios do INSA.
- À Sara Hill e Nuno Faria (Universidade de Oxford) e Joshua Quick e Nick Loman (Universidade de Birmingham) por nos terem gentilmente cedido os primers usados no início deste estudo.
- Às equipas dos projectos [Nextstrain](#) e [Microreact](#) pela libertação de algumas ferramentas de bioinformática usadas neste estudo.
- Ao Miguel Pinheiro (iBiMED / Universidade de Aveiro) pelo seu trabalho na atualização da plataforma [INSaFLU](#) para o novo coronavírus SARS-CoV-2.
- À [Infraestrutura Nacional de Computação Distribuída \(INCD\)](#), por ter fornecido recursos computacionais para testar a plataforma INSaFLU. O INCD foi financiado pela FCT e FEDER sob o projeto 22153-01 / SAICT / 2016.

Este estudo é co-financiado pela Fundação para a Ciência e Tecnologia (234_596874175) no âmbito da "call" Research 4 COVID-19.