



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

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

## Relatório de situação

19 de Maio de 2020

O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **558 sequências do genoma do novo coronavírus SARS-CoV-2**, obtidas de amostras colhidas em 46 laboratórios/hospitais representando 86 concelhos (**Figura 1**).

### Distribuição por "Clade"

- A maioria dos vírus (88.2 %) pertence ao clade A2, sub-clade A2a (**Figura 2**), o mais representado em toda a Europa. O "clade" A2 é caracterizado pela mutação "aa D614G" ("nt A23403G") na proteína "Spike (S)" (**Figure 2**). 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 (6.5 %), A6 (0.5 %), B (3.0 %), outros perfis (1.8 %) (**Figura 2**). O clade A1a apresenta uma dispersão global, com maior frequência no Reino Unido. Paralelamente, em Portugal, foram detectados genomas deste clade em vários Concelhos de Norte a Sul. Relativamente ao clade B, este é maioritariamente composto por genomas detectados em Évora, os quais formam um sub-clade. Na Europa, o clade B apresenta uma maior frequência em Espanha, sugerindo que a sua introdução no Alentejo poderá ter tido origem neste país. Os restantes genomas do clade B formam outro sub-clade e apresentam um perfil genético frequente na Oceânia, sendo congruente com o histórico de viagem de alguns dos casos. Finalmente, o clade A6, frequente nos Países Baixos, foi detectado exclusivamente na Ilha da Madeira, sugerindo uma introdução diferente dos restantes casos observados no arquipélago (ver "destaques" abaixo).

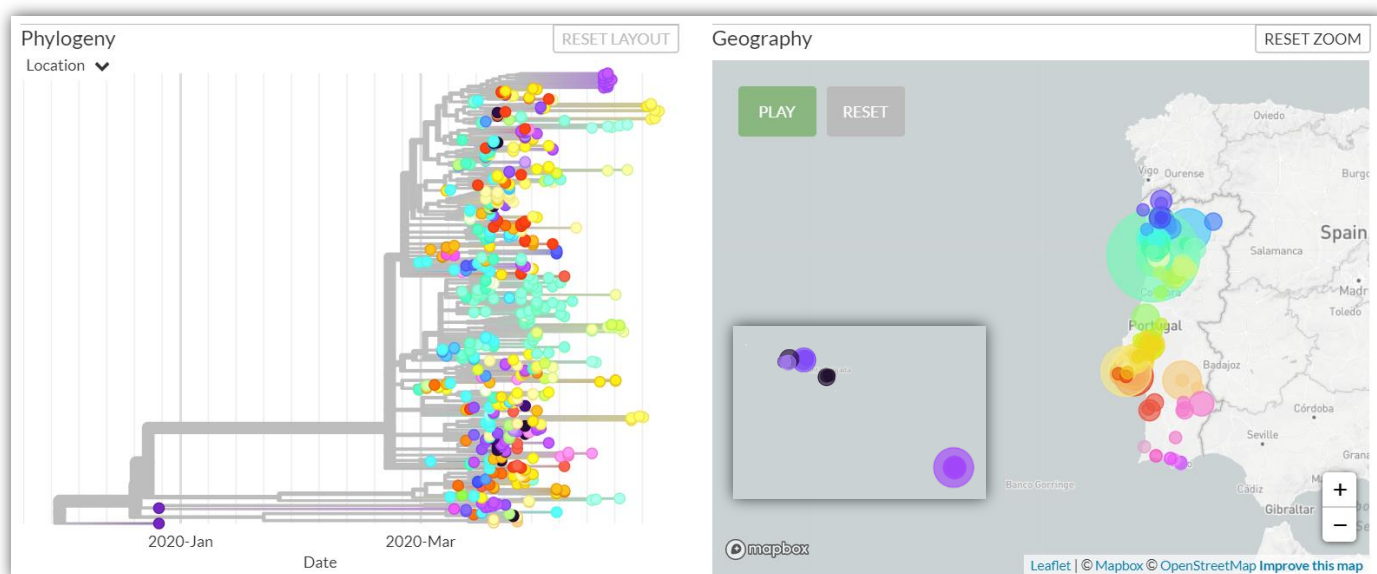
*NOTA: A integração dos dados genómicos de Portugal na diversidade à escala global foi feita com recurso à plataforma Nextstrain (<https://nextstrain.org/ncov>). Nesta actualização, para além da classificação por "Clade" (tal como aplicada na plataforma Nextstrain), foi incluída a classificação por "Linhagem" (Lineage), com base na ferramenta "Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin)" (<https://github.com/hCoV-2019/pangolin>).*

### Mutações – balanço geral

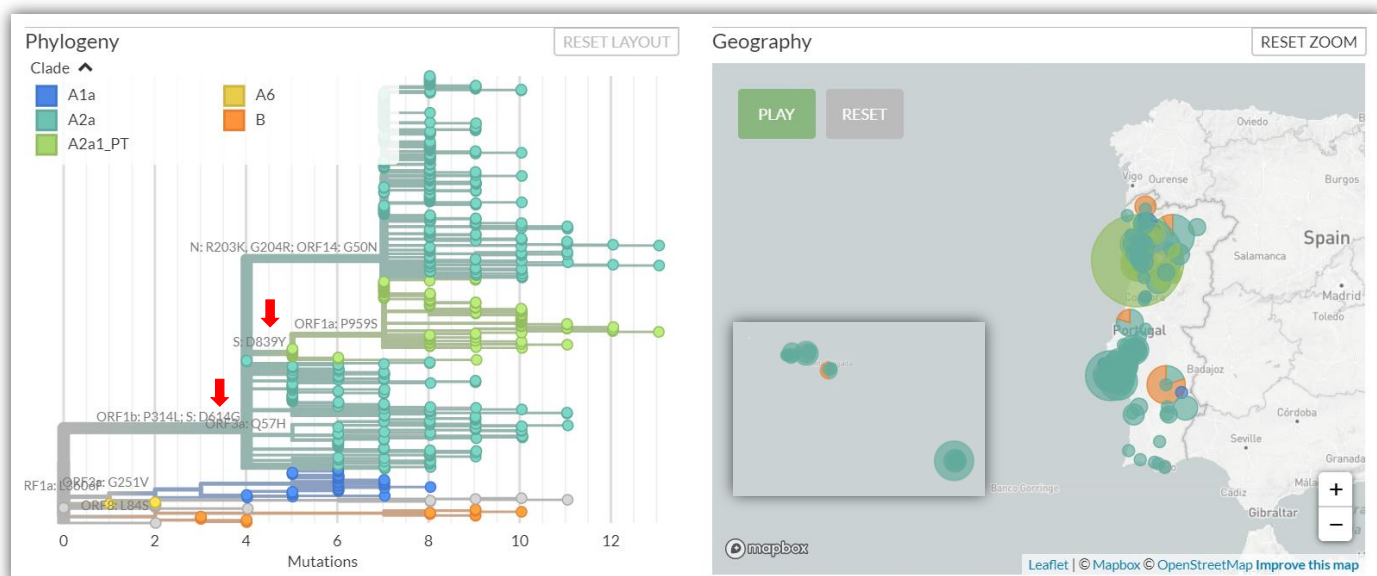
- O número médio de mutações (SNPs) por genoma (comparando com o primeiro genoma sequenciado na China; MN908947.3) é de 8 (variando entre 1 e 13) mutações (**Figura 3A**), o que se enquadra dentro da taxa de mutação prevista para este vírus (i.e., cerca de 2 mutações por genoma por mês). As mutações distribuem-se ao longo dos 29903 nucleótidos que constituem o genoma do novo coronavírus SARS-CoV-2 (**Figura 3B**), tendo sido já detectadas mais de 400 mutações distintas no conjunto dos 558 genomas analisados até à data.
- 40 mutações não-sinónimas alteram a proteína "Spike (S)", sendo que apenas duas delas (D614G e D839Y) estão presentes em mais de 15% dos vírus analisados (88.2% e 18.6%, respectivamente). Curiosamente, enquanto que a mutação D614G caracteriza o clade A2 altamente representado na Europa (**Figura 2**), a mutação D839Y é exclusiva do sub-clade ("A2a1\_PT") que inclui a grande maioria dos genomas detectados em Ovar (**Figura 2**).

### DESTAQUES

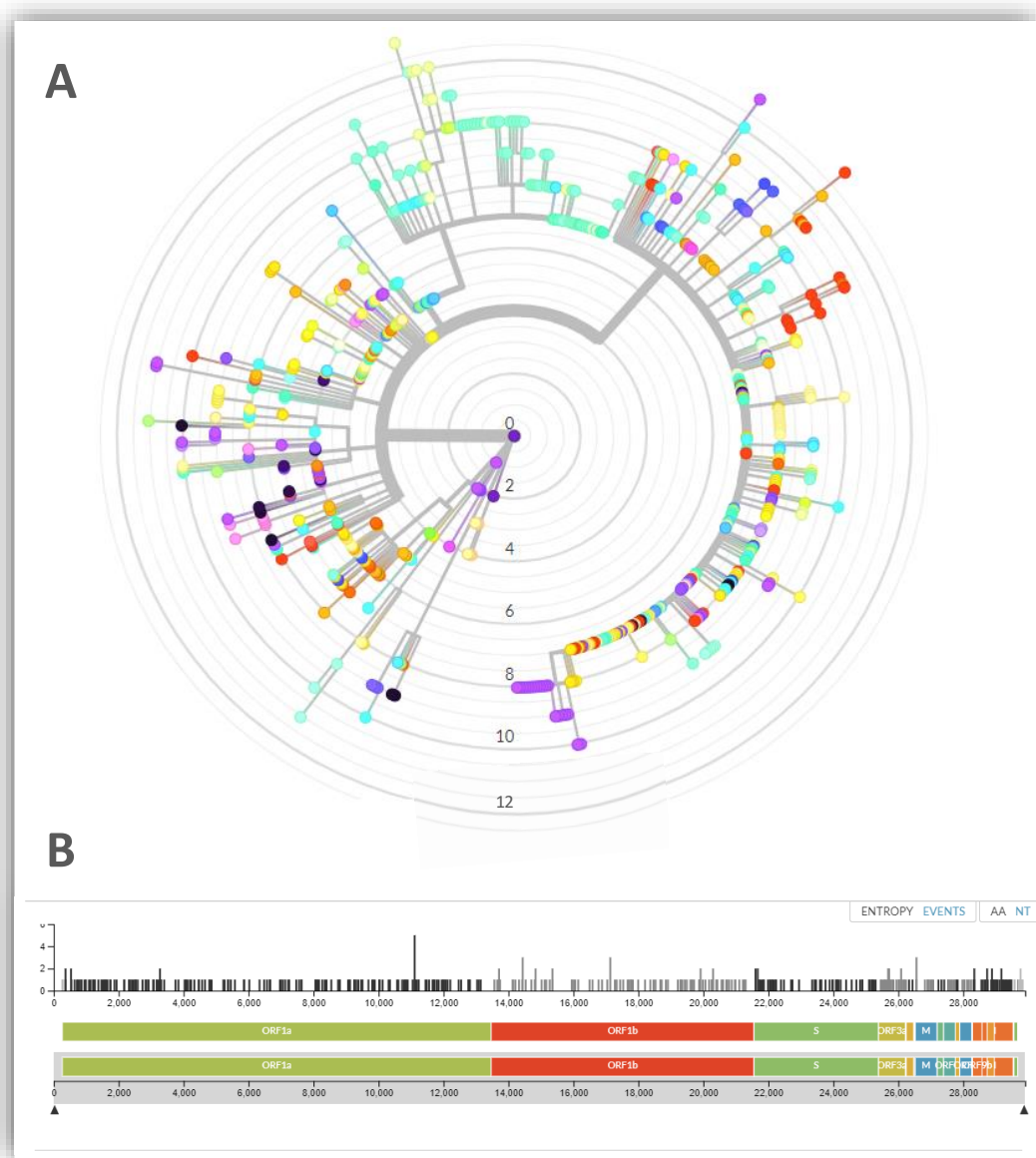
- Observam-se múltiplos clusters filogenéticos congruentes com dados epidemiológicos (ex, geografia, contactos confirmados, etc) e outros sugestivos de potencial link epidemiológico.
- Observa-se um grande cluster ("A2a1\_PT") constituído pela maioria dos genomas avaliados do concelho de Ovar. Este dado sugere 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. 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 (**Figura 4**).
- Analisaram-se até à data mais de 80 amostras provenientes dos arquipélagos. Na Madeira (**Figura 5**), à semelhança dos Açores, detectaram-se genomas pertencentes a vários clades, suportando a existência de distintas introduções do vírus nas ilhas, algumas das quais desencadeando transmissão local.



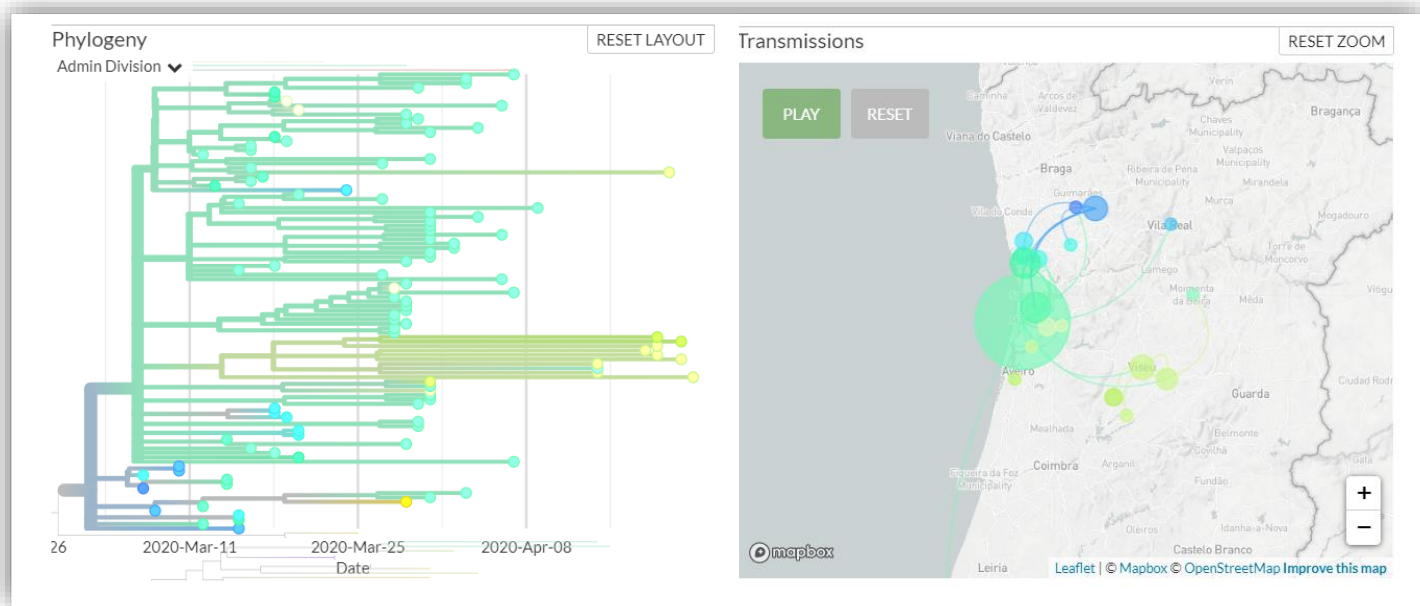
**Figura 1. Visão global da diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal.** Os diferentes genomas (representados por círculos no painel à esquerda) estão coloridos de acordo com o local de residência, com a mesma tonalidade no mapa – o tamanho dos círculos no mapa é proporcional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



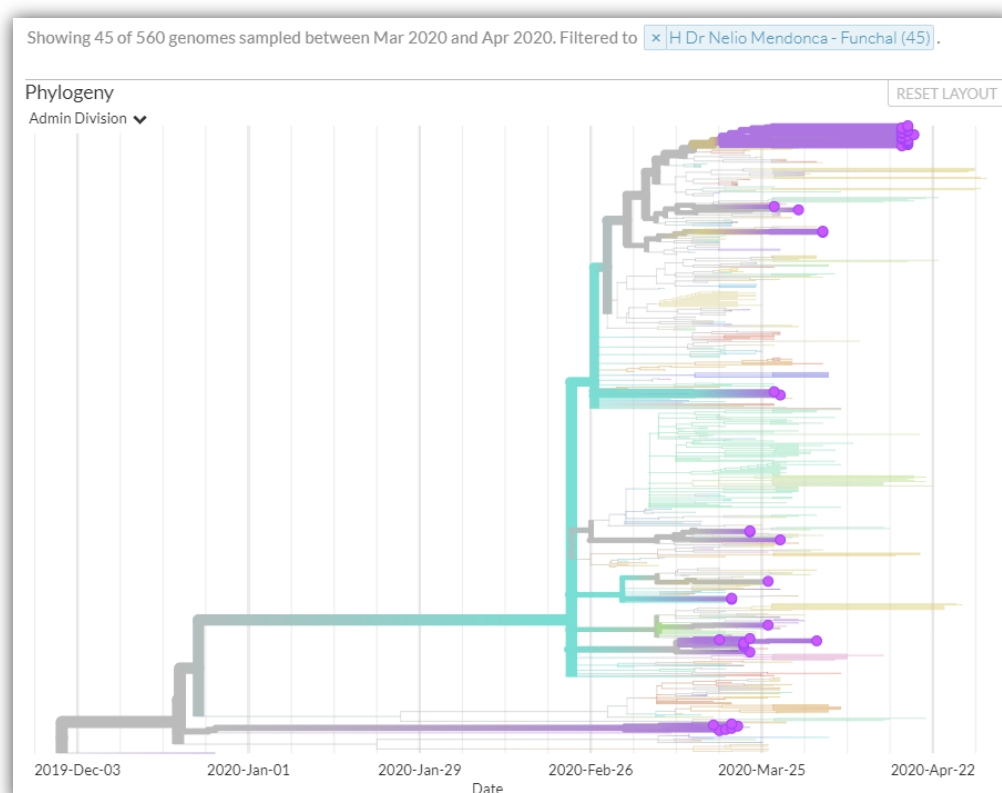
**Figura 2. Diversidade genética e dispersão geográfica do vírus SARS-CoV-2 por “clade” genético, tal como definido em <https://nextstrain.org/>, adicionando o sub-clade A2a1\_PT relativo à potencial cadeia de transmissão mais evidente até agora, com foco em Ovar. As setas a vermelho sinalizam as mutações **D614G** e **D839Y** na proteína “Spike (S)”, as quais caracterizam o clade **A2/A2a** (altamente representado na Europa) e o sub-clade **A2a1\_PT** (frequente em algumas regiões de Portugal), respectivamente.**



**Figura 3. A) 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 de acordo com o concelho de residência). **B) 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**.



**Figura 4. Diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 pertencente ao sub-clade “A2a1\_PT”.** Este sub-clade, o qual tem como “marcador” a mutação D839Y na proteína “Spike (S)”, concentra a maioria dos genomas avaliados do concelho de Ovar. O perfil genético e de disseminação geotemporal aponta que este foco em **Ovar** 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.



**Figura 5. Diversidade genética e dispersão geotemporal dos vírus SARS-CoV-2 detectados na Madeira (coloridos a roxo).** Neste arquipélago observam-se genomas pertencentes a vários clades, suportando a existência de distintas introduções do vírus, algumas das quais desencadeando transmissão local (p.ex., cluster no topo da árvore).



## 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, bem como **monitorização das mutações em alvos genéticos de testes de diagnóstico**.
- 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-bbmuk6w>)
- **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 em <https://nextstrain.org/ncov> e <https://github.com/nextstrain/ncov>) e [Microreact](#).

\* 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.**