



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

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

## Relatório de situação

24 de Junho de 2020

O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **976 sequências do genoma do novo coronavírus SARS-CoV-2**, obtidas de amostras colhidas em 51 laboratórios/hospitais/instituições representando 131 concelhos (Figura 1).

### Distribuição por "Clade"

- A distribuição por clade (Tabela 1) é, em termos gerais, semelhante àquela que é observada a nível Europeu (<https://nextstrain.org/ncov/europe>). A maioria dos vírus (89%) integra o braço filogenético contendo os clades 20A, 20B e 20C, os quais apresentam, entre outros marcadores genéticos, a mutação "aa D614G" ("nt A23403G") na proteína "Spike (S)". Esta proteína é responsável pela entrada do vírus SARS-CoV-2 nas células humanas, sendo também o principal antígeno deste vírus pandémico.

Tabela 1. Distribuição dos genomas por "Clade"

Clade	n.º de genomas	%
19A	77	8%
19B	31	3%
20A	407	42%
20B	437	45%
20C	24	2%

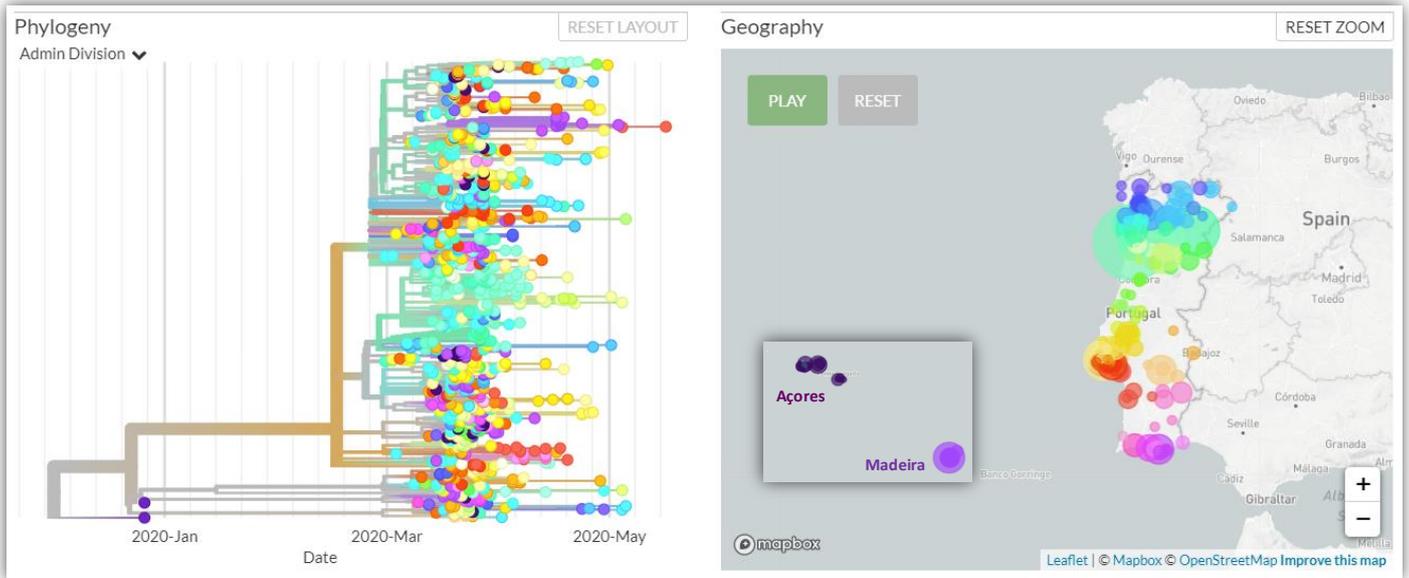
- Nesta actualização do [site](#) foi integrada a possibilidade de verificar, de um modo simples, a **distribuição dos diferentes Clades por Distrito e por Concelho (Figura 2)**, usando a plataforma [Microreact](#) (Nota: estes links dão também acesso a um filme que mostra a distribuição filogeográfica ao longo do tempo). Desta análise, destaca-se, por exemplo, a observação de que **os vírus do clade 19B têm sido detectados maioritariamente em concelhos próximos da fronteira com Espanha** (por exemplo, Arcos de Valdevez, Bragança, Miranda do Douro ou Évora). Este dado é consistente com a elevada circulação de SARS-CoV-2 com perfil 19B em Espanha (ao contrário daquilo que se observa na maioria dos países Europeus) e sugere que **estas introduções ocorreram através da fronteira terrestre**. Não obstante, foi já possível detectar introduções independentes de vírus 19B com associação a outras regiões geográficas onde este perfil também é observado (por exemplo, casos nos Açores com história de viagem à Oceânia).

### Mutações – balanço geral

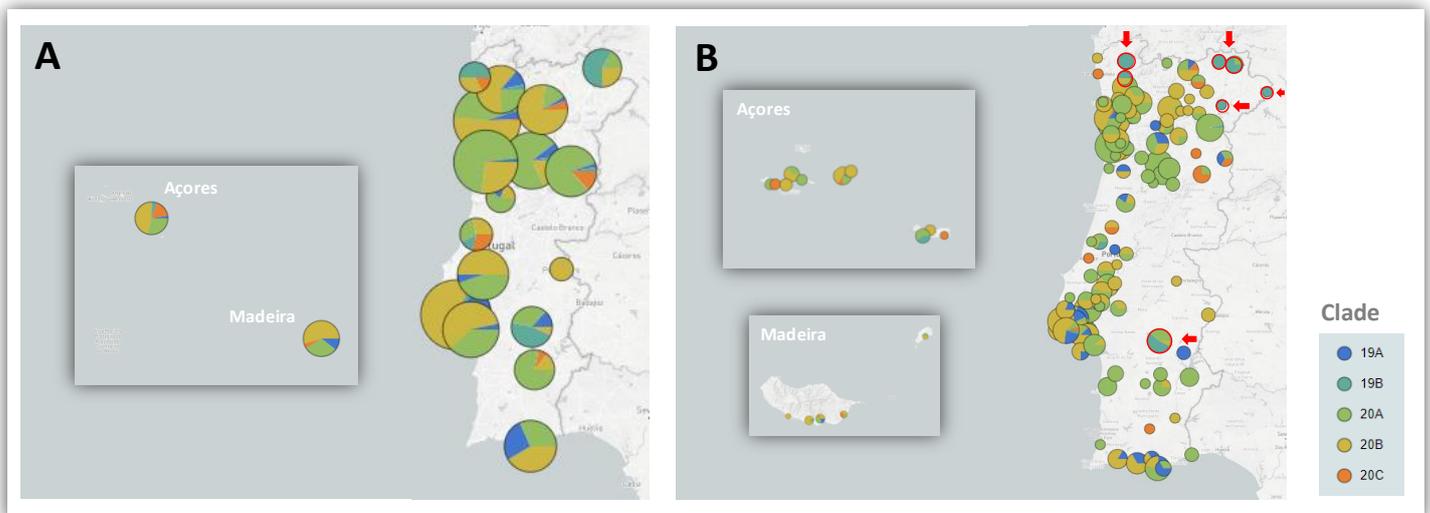
- O **número médio de mutações (SNPs) por genoma** (comparando com o primeiro genoma sequenciado na China; MN908947.3) é de **8 (variando entre 1 e 20) mutações**, o que se enquadra dentro da taxa de mutação prevista para este vírus (i.e., cerca de 2 mutações por genoma por mês). No conjunto dos 976 genomas analisados até à data, observam-se SNPs em 724 dos 29903 nucleótidos que constituem o genoma do novo coronavírus SARS-CoV-2, sendo que **63 são SNPs não-sinónimas que alteram a proteína "Spike (S)"**.
- Para além de duas mutações "prevalentes" na proteína S (ver detalhes no relatório de 19 de Maio), **D614G (altamente prevalente à escala europeia e nacional) e D839Y** (marcadora do sub-clade que inclui a grande **maioria dos genomas detectados em Ovar e Distrito de Viseu**), destaca-se, neste relatório, a detecção (em apenas **um genoma**) de uma mutação (**P491S**) na **região de ligação da proteína "Spike (S)" ao seu receptor (ACE2) das células humanas**. Esta mutação **não tinha sido reportada a nível mundial** (<https://nextstrain.org/ncov/global>). Actualmente, não se conhece o impacto destas mutações na capacidade do SARS-CoV-2 se transmitir ou causar doença.

### OUTROS DESTAQUES

- Foi possível confirmar que um **importante foco de infecção reportado em Vila Nova de Foz Côa no final de Março** foi causado por SARS-CoV-2 pertencente ao **sub-clade com a mutação marcadora D839Y associado ao surto em Ovar (Figura 3)**. Esta mutação é raramente encontrada em outros países, contudo o primeiro genoma reportado com este perfil foi detectado em Itália (região da Lombardia) (Figura 4), sugerindo que a sua introdução em Portugal poderá ter tido essa origem. O INSA está a trabalhar de forma estreita com as autoridades de Saúde locais por forma a desvendar quando e como é que este perfil foi introduzido em Portugal, a sua dinâmica de disseminação e potenciais associações com dados clínicos.
- Na última actualização do [site](#) (dia 23/06/2020) **foram incluídas 35 sequências do genoma obtidas no Instituto de Gulbenkian de Ciência (IGC)**, instituição colaboradora do INSA neste estudo.

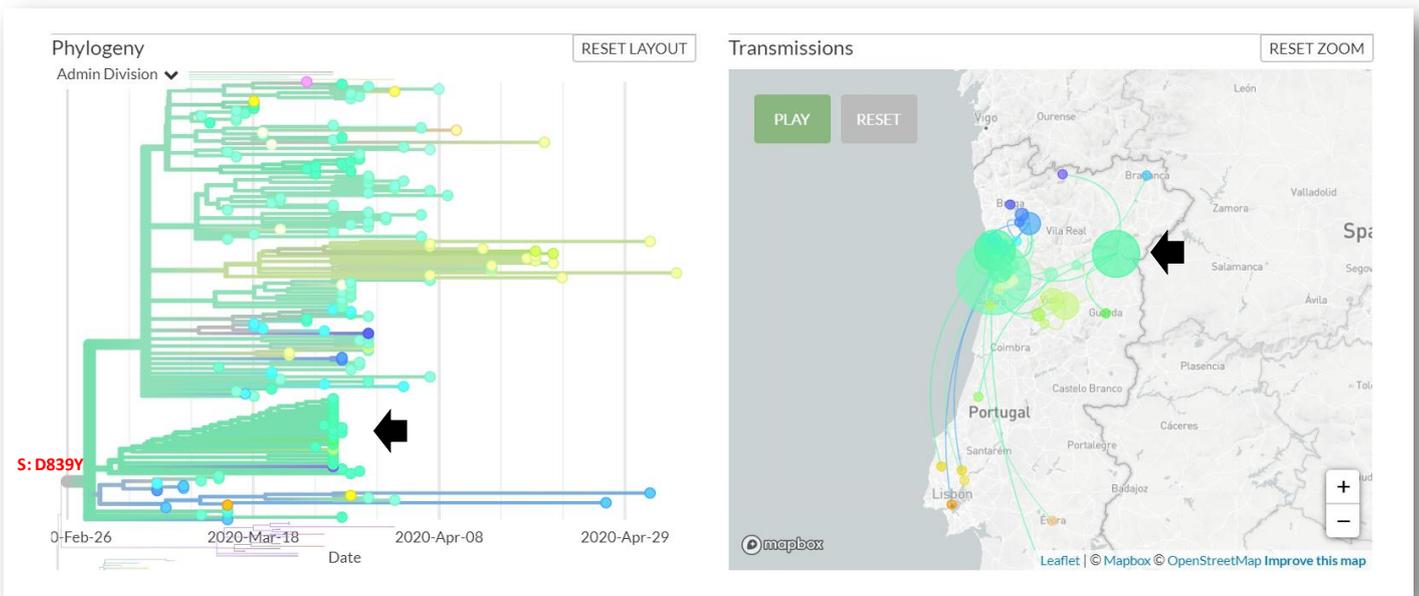


**Figura 1. Visão global da diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal.** Os diferentes genomas (representados por círculos no painel à esquerda) estão coloridos de acordo com o local de residência, com a mesma tonalidade no mapa – o tamanho dos círculos no mapa é proporcional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

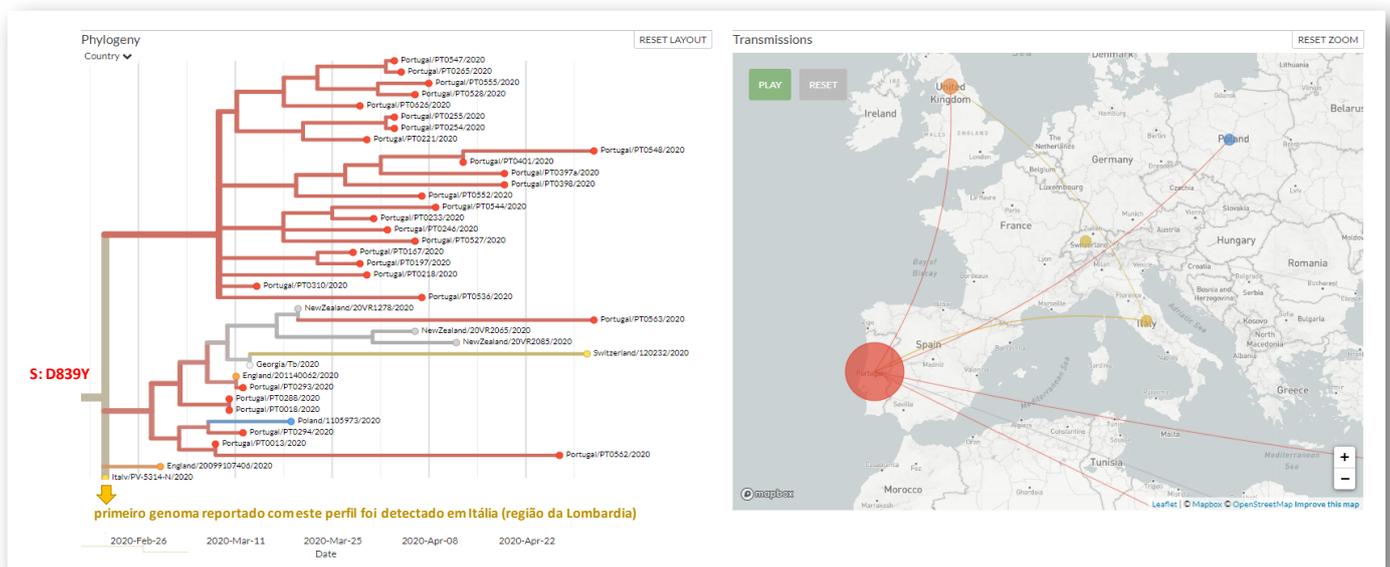


**Figura 2. Dispersão geográfica do vírus SARS-CoV-2 por “clade” genético, com resolução geográfica por: A) Distrito (Portugal Continental) e Arquipélagos; B) Concelho.** No painel B estão destacados, com círculos com contorno vermelho, concelhos em que grande parte dos genomas analisados pertence ao clade 19B, o que é sugestivo da introdução (setas vermelhas) deste perfil através da fronteira terrestre. De facto, na Europa, os vírus com perfil 19B têm sido maioritariamente detectados em Espanha.

No site do [INSA](https://insaflu.insa.pt/), os genomas são classificados por “Clade” adoptando a nomenclatura usada na plataforma “[NextStrain](https://nextstrain.org/)”, estando também disponível a classificação por “Lineage” usando a ferramenta “[Phylogenetic Assignment of Named Global Outbreak Lineages](https://pubmed.ncbi.nlm.nih.gov/34811111/)”, [Rambaut et al. 2020](https://pubmed.ncbi.nlm.nih.gov/34811111/)).



**Figura 3. Diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 pertencente ao sub-clade com a mutação marcadora D839Y na proteína “Spike (S)”.** Este sub-clade concentra a maioria dos genomas avaliados do Concelho de Ovar e do Distrito de Viseu. Na última actualização do site, foi possível confirmar que um importante foco de infecção (sinalizado com setas pretas) reportado em Vila Nova de Foz Côa no final de Março foi também causado por SARS-CoV-2 pertencente a este sub-clade.



**Figura 4. Enquadramento do sub-clade com a mutação marcadora D839Y na proteína “Spike (S)” no contexto mundial** (genomas coloridos de acordo com o país). Virus deste sub-clade foram detectados pela primeira vez em Itália (região da Lombardia) no final de Fevereiro de 2020. Este perfil teve grande “expressão” em Portugal (estando fortemente associado às regiões de Ovar, Vila Nova de Foz Côa e Viseu), ao contrário da sua aparente limitada disseminação noutras partes do mundo, incluindo Itália.

NOTA: Imagem de <https://nextstrain.org/ncov/europe> (reflecta uma sub-amostragem dos genomas pertencentes a este sub-clade).



## 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 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 em 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.**