



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

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

## Relatório de situação

4 de Junho de 2020

O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **781 sequências do genoma do novo coronavírus SARS-CoV-2**, obtidas de amostras colhidas em 47 laboratórios/hospitais/instituições representando 116 concelhos (**Figura 1**).

### Distribuição por “Clade” (nova classificação)

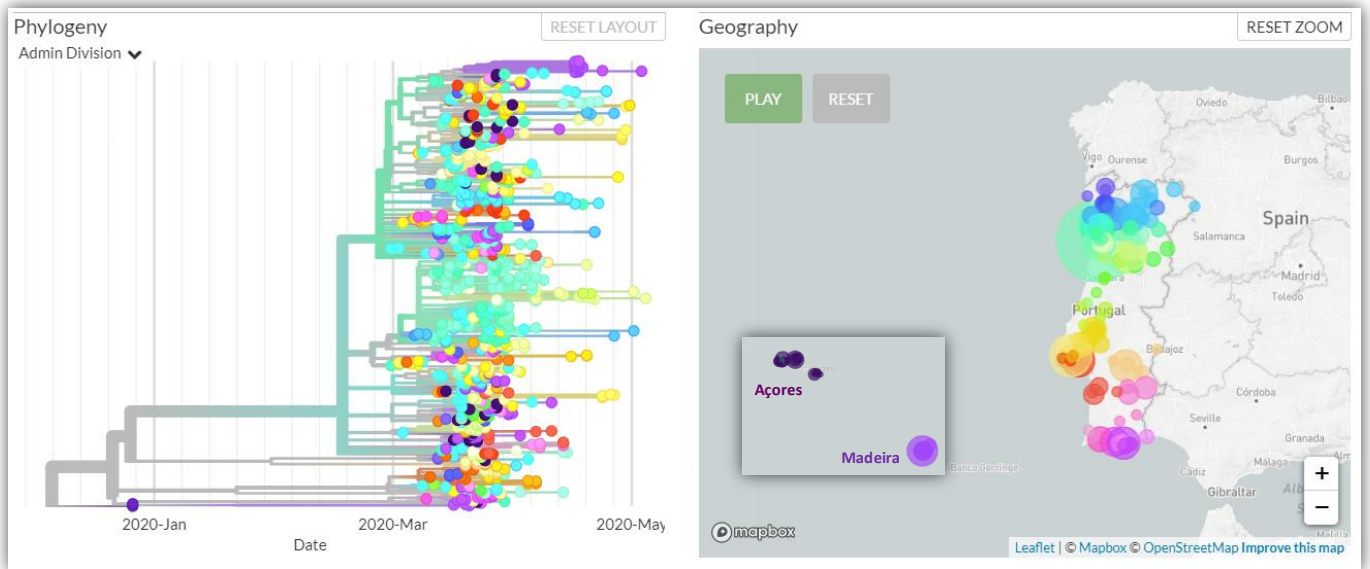
- Os vírus que circulam em Portugal apresentam uma distribuição, por clade, semelhante à que é observada a nível Europeu. A maioria dos vírus (88.6%) pertence aos clades **20A (41.5%), 20B (44.4%) e 20C (2.7%) que são caracterizados pela mutação “aa D614G” (“nt A23403G”) na proteína “Spike (S)” (Figura 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. De notar que neste relatório se adoptou a nova classificação “NextStrain” (<https://nextstrain.org/ncov/europe>).
- Entre os restantes clades, o clade **19B (2.6%) apresenta um sub-clade** maioritariamente composto por genomas detectados em Évora. Na Europa, este sub-clade apresenta uma maior frequência em Espanha, sugerindo que a sua introdução naquela região do Alentejo poderá ter tido origem neste país. Os restantes genomas do clade 19B formam outro sub-clade e apresentam um perfil genético frequente na Oceânia, sendo congruente com o histórico de viagem de alguns dos casos. (**Figura 2**). O clade **19A (8.8%)**, ao qual pertence o primeiro genoma de SARS-CoV-2 obtido na China, apresenta uma dispersão global. Paralelamente, em Portugal, foram detectados genomas deste clade em vários Concelhos de Norte a Sul.

### Mutações – balanço geral

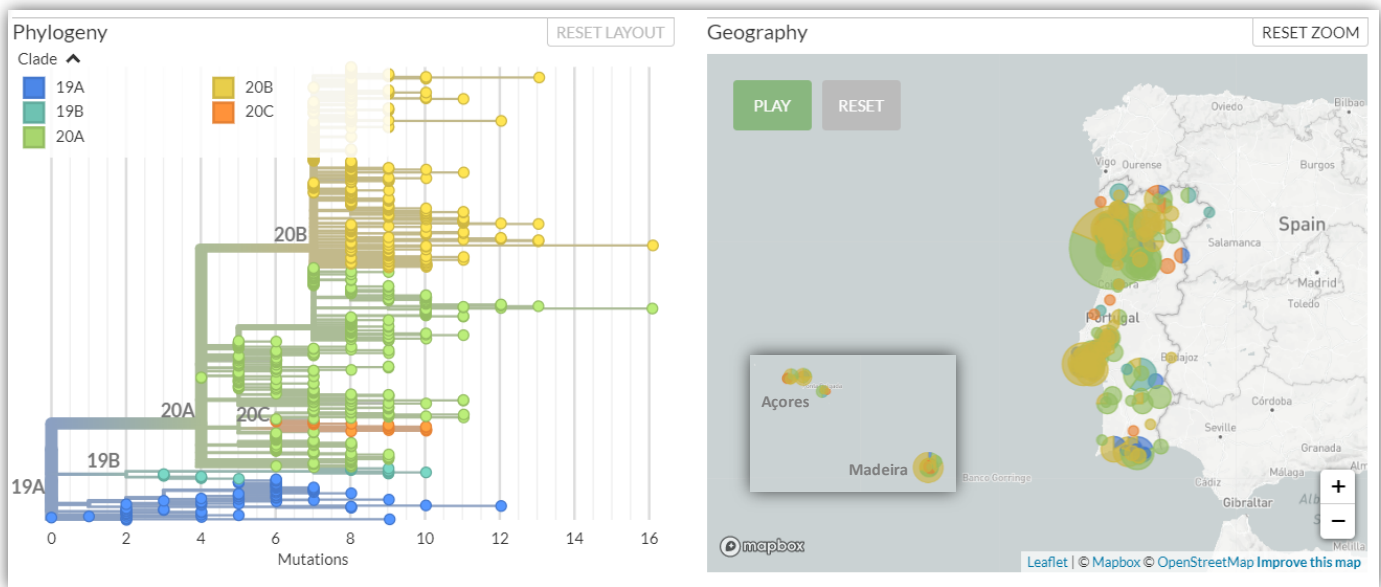
- O número médio de mutações (SNPs) por genoma (comparando com o primeiro genoma sequenciado na China; MN908947.3) é de **8 (variando entre 1 e 16) mutações (Figura 3A)**, o que se enquadra dentro da taxa de mutação prevista para este vírus (i.e., cerca de 2 mutações por genoma por mês). As mutações distribuem-se ao longo dos 29903 nucleótidos que constituem o genoma do novo coronavírus SARS-CoV-2 (**Figura 3B**), tendo sido já detectadas perto de 600 mutações distintas no conjunto dos 781 genomas analisados até à data.
- 47 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 (88.6% e 19.0%, respectivamente). Curiosamente, ao passo que a mutação **D614G caracteriza os vírus dos clades 20A, 20B e 20C, altamente representados na Europa (Figura 2)**, a mutação **D839Y** é exclusiva do sub-clade que inclui a grande maioria dos genomas detectados em Ovar (ver detalhes no relatório de 19 de Maio).

### DESTAQUES

- O Instituto Ricardo Jorge está a colaborar activamente com múltiplas Autoridades de Saúde Locais de forma a integrar dados epidemiológicos e genómicos e, assim, **perceber, de forma mais robusta, os pontos de entrada e cadeias de transmissão a nível regional e local**. Destacam-se, por exemplo, as colaborações com as equipas de Saúde Pública do arquipélago da Madeira e da região de Ovar, cujos resultados da análise do genoma revelaram já importantes pistas neste âmbito (ver detalhes no relatório de 19 de Maio). No caso da Madeira, espera-se que estes resultados, já em fase de análise, sejam ultimados em breve.
- Neste relatório destacamos as **duas primeiras semanas da epidemia em Portugal**, desde a detecção dos primeiros casos a 2 de Março até dia 15 de Março, altura em que as autoridades de saúde monitorizavam já **18 cadeias de transmissão ativa (DGS, 16 de Março de 2020)**. Os dados genómicos dos vírus deste período (**Figura 4**) revelam que até 15 de Março foi introduzida uma ampla diversidade genética (tendo-se detectado já vírus representativos de todos os clades) concordante com um número bastante mais elevado de introduções não relacionadas do vírus em Portugal. A monitorização destas introduções, bem como a confirmação de que estas terão, ou não, desencadeado cadeias de transmissão de menor ou maior magnitude (como no caso de Ovar), será continuamente revisitada à medida que se adquirem mais dados.

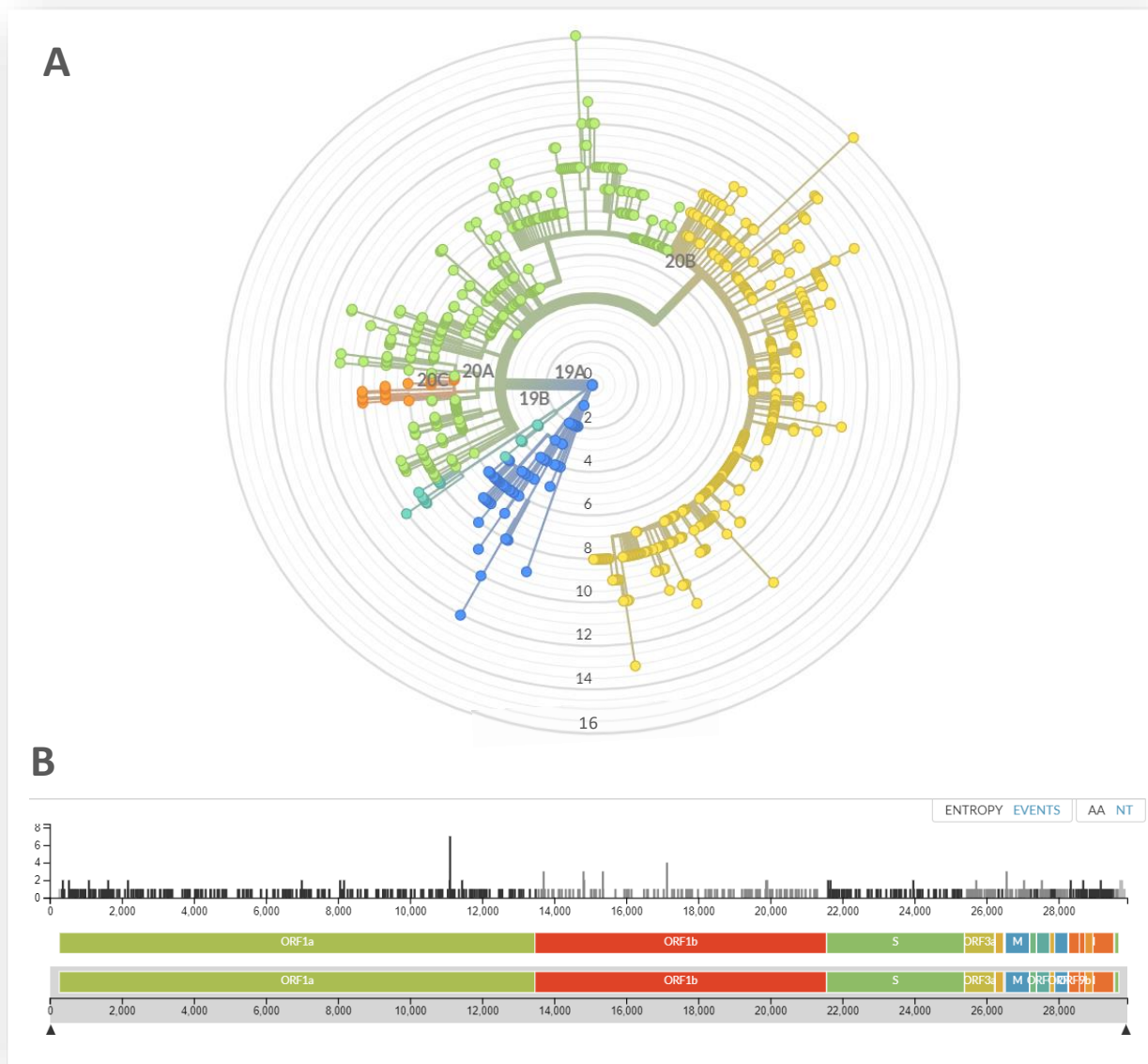


**Figura 1. Visão global da diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal.** Os diferentes genomas (representados por círculos no painel à esquerda) estão coloridos de acordo com o local de residência, com a mesma tonalidade no mapa – o tamanho dos círculos no mapa é proporcional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

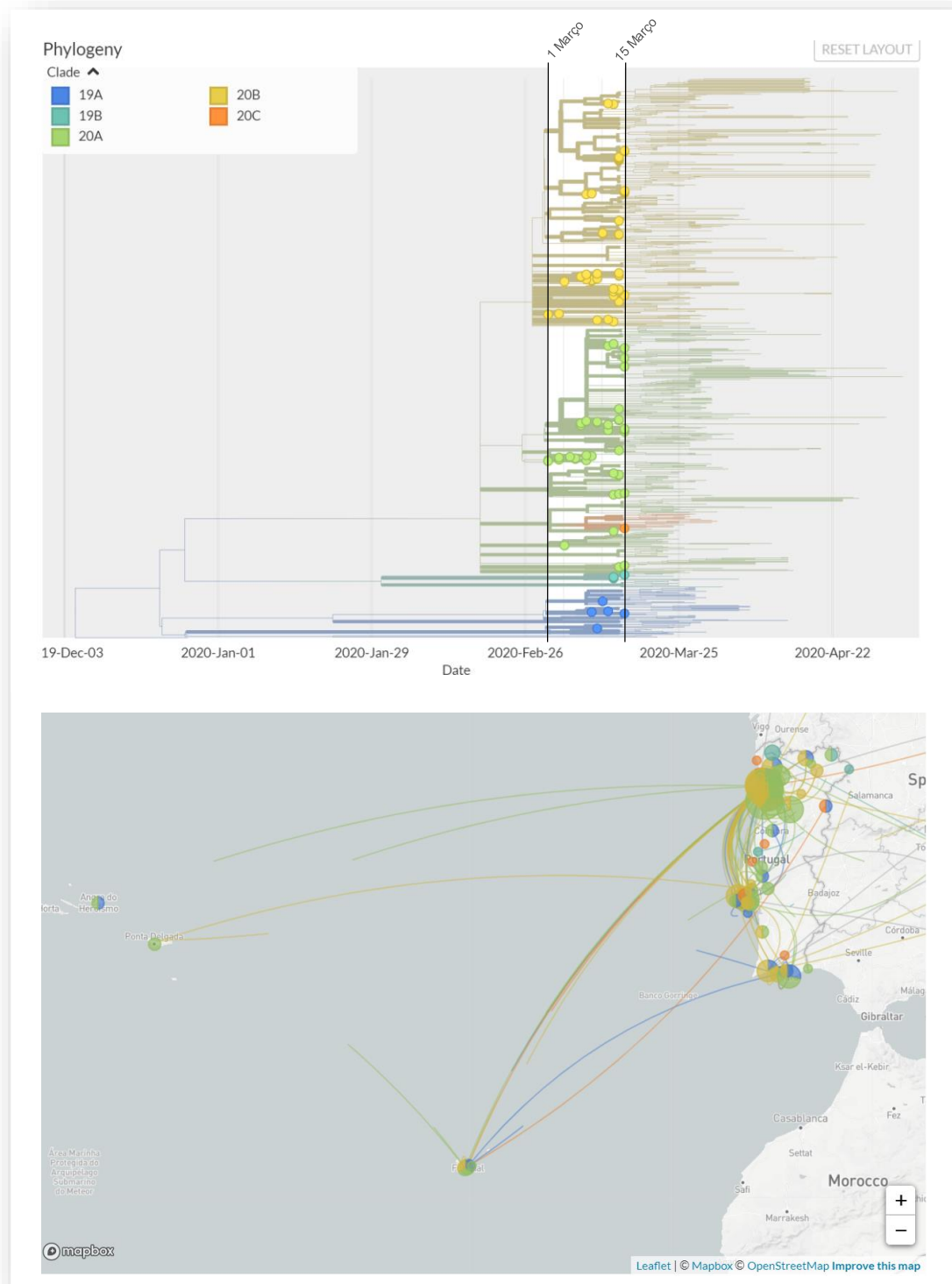


**Figura 2. Diversidade genética e dispersão geográfica do vírus SARS-CoV-2 por “clade” genético, tal como definido em <https://nextstrain.org/>.**

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



**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 clade genético). **B) Representação gráfica do genoma do novo coronavírus SARS-CoV-2** (com diferentes cores para diferentes genes), em que, imediatamente 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 dos vírus SARS-CoV-2 (coloridos por clade genético) detectados nas duas primeiras semanas da epidemia em Portugal (1 a 15 de Março de 2020).** Os vírus detectados neste período pertenciam já aos clades genéticos actualmente em circulação, apontado para um cenário de introduções múltiplas e não relacionadas do vírus em várias regiões de Portugal no final de Fevereiro/início de Março. A adopção de medidas de restrição de entrada de pessoas em Portugal ocorreu maioritariamente no final ou após este período (e.g., fecho fronteiras terrestres, forte limitação de voos, nomeadamente de e para Espanha e Itália).



**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 antigénios 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 <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 / SACT / 2016.

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