



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

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

Relatório de situação

23 de Dezembro de 2020

O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **2234** sequências do genoma do novo coronavírus SARS-CoV-2, obtidas de amostras colhidas em 68 laboratórios/hospitais/instituições representando 199 concelhos (Figura 1). Nesta actualização do site foram inseridas **mais 449** sequências do genoma de SARS-CoV-2, sendo que 196 foram geradas no Instituto de Gulbenkian de Ciência (IGC), instituição colaboradora do INSA neste estudo.

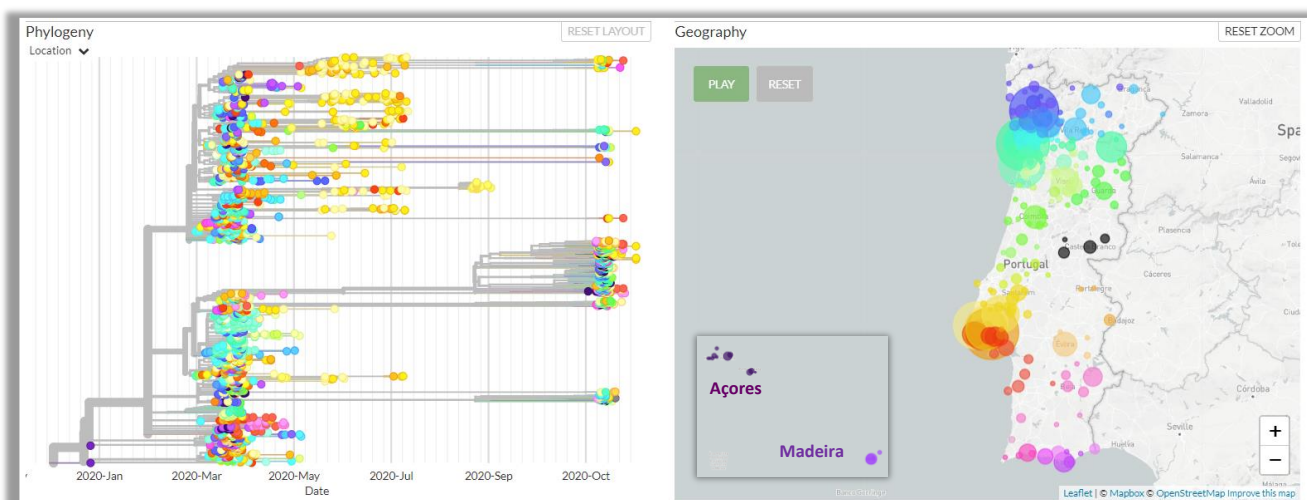


Figura 1. Visão global da diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal. Os diferentes genomas (representados por círculos no painel à esquerda) estão coloridos de acordo com o local de residência / colheita, com a mesma tonalidade no mapa – o tamanho dos círculos no mapa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

Entre as novas sequências, destacam-se **402** sequências obtidas de amostras colhidas desde o início de Novembro (Figura 2) em 19 laboratórios/hospitais colaboradores dispersos por todo território nacional (estão representados 16 distritos de Portugal continental e o Arquipélago dos Açores, num total de 113 concelhos).

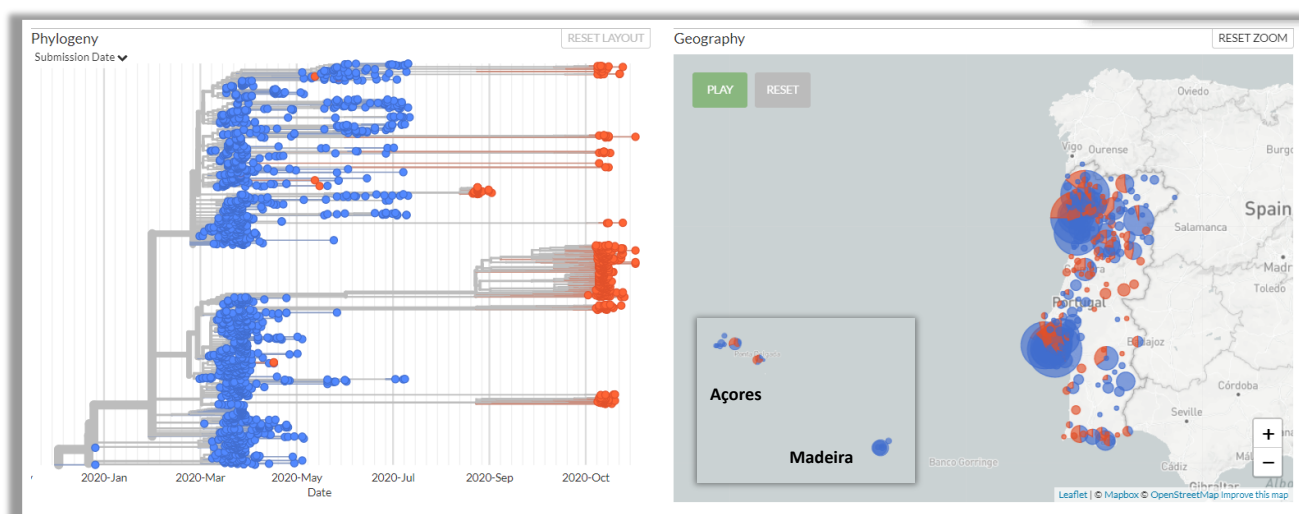


Figura 2. Enquadramento das novas 402 sequências obtidas desde Novembro (círculos a laranja após 1 de Novembro) na diversidade genética e dispersão geotemporal do vírus SARS-CoV-2 em Portugal. O tamanho dos círculos no mapa é proporcional ao número de genomas sequenciados por concelho durante este período.



O presente relatório pretende resumir a distribuição e frequência das variantes genéticas do SARS-CoV-2 detectadas em Portugal desde o início de Novembro.

Variantes da proteína Spike

- Todos os vírus analisados desde Novembro (n=402) integram 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)". Entre os 402 vírus sequenciados foram detectadas 78 diferentes alterações aminoacídicas na proteína Spike, sendo que algumas são marcadoras de clades sob vigilância a nível internacional. **A Tabela resume os clades mais relevantes (Figura 3), bem como dados epidemiológicos / biológicos acerca desses clades e das suas mutações marcadoras.**

Variante	PROPORÇÃO (n=402)	CONTEXTO
S: D614G+A222V (sub-clade "20A.EU1")	71,9%	- Este cluster (sub-clade 20A.EU1) terá tido origem em Espanha (em meados de Junho), tendo revelado uma elevada disseminação a nível global. Hodcroft et, al, medRxiv. https://www.medrxiv.org/content/10.1101/2020.10.25.20219063v2 - A mutação A222V na Spike altera um potencial local antigénico (epítipo). Mateus et al, 2020, Science. https://science.sciencemag.org/content/370/6512/89 - Uma variante com esta mutação A222V este na origem de um caso de re-infecção descrito em Hong Kong. To et al, 2020, CID. https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa1275/5897019
S: D614G+S477N (sub-clade "20A.EU2")	9,0%	- Este cluster (sub-clade 20A.EU2) apresenta uma considerável disseminação internacional. Hodcroft et, al, medRxiv. https://www.medrxiv.org/content/10.1101/2020.10.25.20219063v2 - A mutação S477N na Spike altera o domínio de ligação ao receptor ("RBD") ACE2, podendo aumentar a afinidade da ligação da Spike às células do hospedeiro. Starr et al, 2020, Cell. https://www.sciencedirect.com/science/article/pii/S0092867420310035
S: D614G+S98F (sub-clade 20A)	4,7%	- Esta variante é encontrada com maior frequência na Bélgica e Holanda, tendo já sido identificada em múltiplos países. Hodcroft et, al, medRxiv. https://www.medrxiv.org/content/10.1101/2020.10.25.20219063v2 ; https://github.com/emmahodcroft/cluster_scripts/blob/master/README.md#sn501
S: D614G+D80Y (sub-clade 20C)	1,0%	- Esta variante é encontrada com maior frequência na França, Noruega e Holanda, tendo já sido identificada em múltiplos países. Hodcroft et, al, medRxiv. https://www.medrxiv.org/content/10.1101/2020.10.25.20219063v2 ; https://github.com/emmahodcroft/cluster_scripts/blob/master/README.md#sn501
OUTROS	13,4%	----



- **As 3 variantes mais frequentes** em Novembro (**Figura 3**), cada uma delas reconhecida por uma alteração diferente na proteína Spike (**A222V**, **S477N** ou **S98F**), foram detectadas em todas as regiões de Portugal continental, sugerindo que **terão sido as principais co-responsáveis pela segunda vaga epidémica de SARS-CoV-2**.

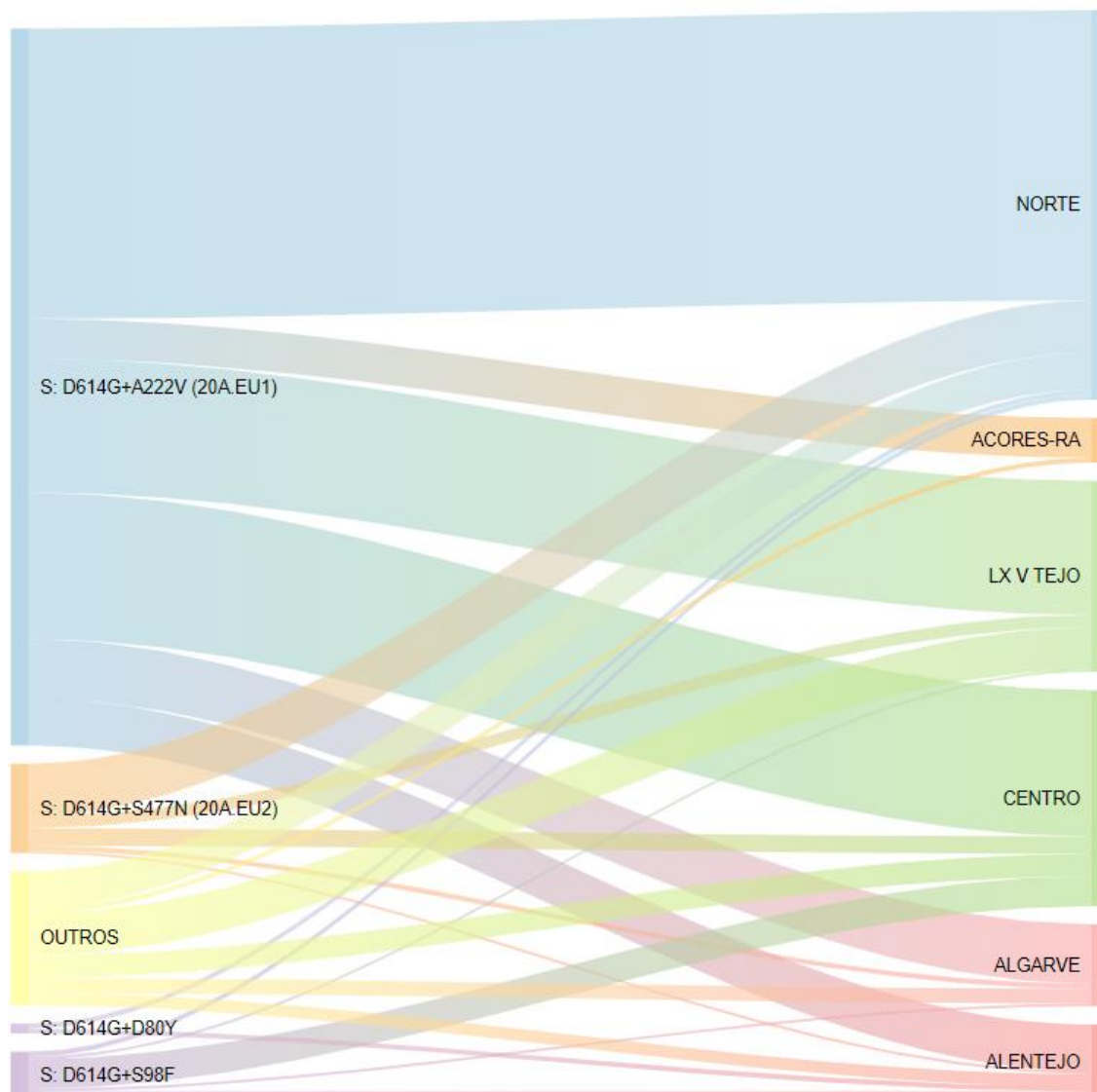


Figura 3. Frequência das principais variantes genéticas do SARS-CoV-2 detectadas em Portugal desde o início de Novembro (à esquerda) e sua distribuição por região (à direita).



- A variante **D614G+A222V** (sub-clade 20A.EU1) destaca-se pela sua elevada frequência (289/402 novas sequências) e ampla distribuição geográfica. De facto, foi detectada em todos os distritos cobertos nesta amostragem (Figura 4). Dentro deste sub-clade 20A.EU1 detectaram-se vírus com mutações adicionais na Spike, destacando-se as alterações P1162R (presente em 80/289 vírus deste grupo), a combinação A262S+P272L (presente em 7 vírus, um dos quais apresentando a alteração R458K no domínio de ligação ao receptor das células humanas).

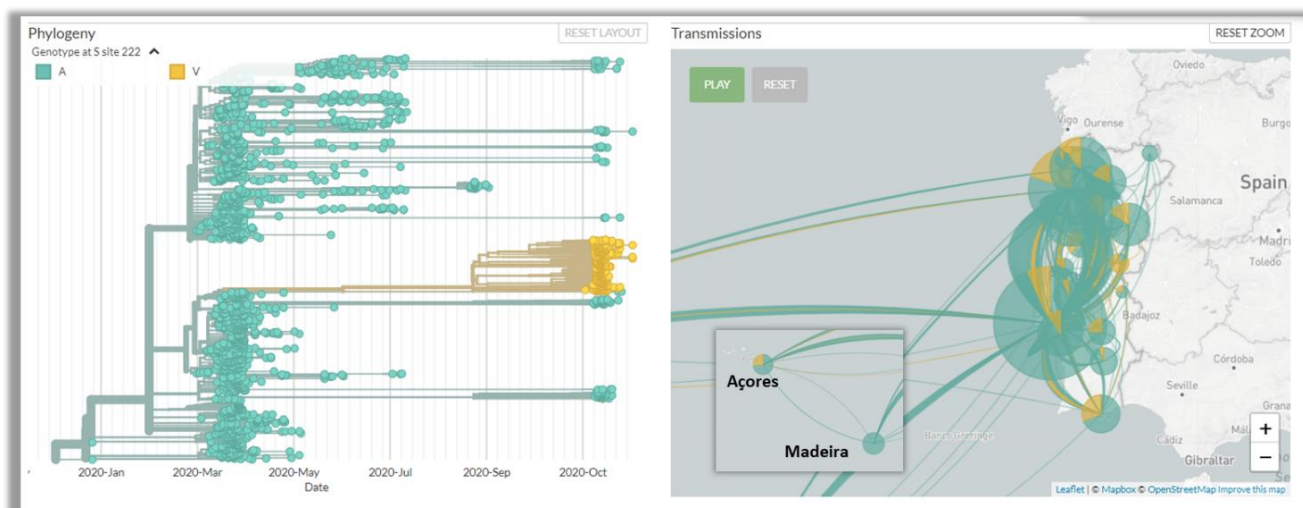


Figura 4. Enquadramento e dispersão geográfica do sub-clade 20A.EU1 (estes vírus partilham a combinação de mutações **D614G + A222V**, na proteína Spike) no contexto do global dos genomas analisados em Portugal. Os genomas com e sem a mutação A222V estão coloridos a amarelo e verde, respetivamente, sendo o tamanho dos círculos no mapa proporcional ao número de genomas sequenciados por localidade.

OUTROS DESTAQUES

Variantes não detectadas:

- Entre os 402 vírus sequenciados recentemente **não se encontrou qualquer vírus com a combinação de mutações apresentada pela variante (pertencente à linhagem B.1.1.7) actualmente a circular no Reino Unido¹ ou pelas variantes associadas a surtos de SARS-CoV-2 em explorações de martas na Dinamarca².**
- De facto, entre as múltiplas mutações na Spike que têm merecido atenção das autoridades de saúde a nível internacional, não foi detectado qualquer SARS-CoV-2 com as seguintes mutações de especial enfoque na actualidade: N501Y¹ (associada às novas variantes actualmente a circular no Reino Unido e na África do Sul); **Y453F²** (reportada em casos de infecção em animais, tais como martas/visons); **H69del/V70del²** (associada à nova variante a circular no Reino Unido e a casos de infecção em animais, tais como martas/visons), e **N439Y²** (potencialmente mediadora de resistência a anticorpos neutralizantes).
- Curiosamente, **não foi detectada a variante genética que marcou a primeira vaga da epidemia de COVID-19 em Portugal** (com a mutação na Spike **D839Y**; ver relatório de 21 de agosto) nos genomas analisados da segunda vaga. Isto sugere que medidas de saúde pública implementadas terão mitigado a continuação da sua transmissão.

Referências:

1. European Centre for Disease Prevention and Control. Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom – 20 December 2020. ECDC: Stockholm; 2020; <https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf>
2. European Centre for Disease Prevention and Control. Detection of new SARS-CoV-2 variants related to mink – 12 November 2020. ECDC: Stockholm; 2020; <https://www.ecdc.europa.eu/sites/default/files/documents/RRA-SARS-CoV-2-in-mink-12-nov-2020.pdf>



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-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 (Distrito – "Admin Division" ou Concelho – "Location") de residência ou, caso não exista informação, local de ocorrência ou da entidade que enviou a amostra.

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 e Agência de Investigação Clínica e Inovação Biomédica no âmbito da primeira e segunda "calls" do programa Research 4 COVID-19 (projectos 234 e 657, respectivamente).