



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

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

Relatório de situação

28 de Dezembro de 2020

O Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P. (INSA) analisou até à data **2268 sequências do genoma do novo coronavírus SARS-CoV-2**, obtidas de amostras colhidas em 68 laboratórios/hospitais/instituições representando 199 concelhos. 593 sequências foram geradas no Instituto de Gulbenkian de Ciência (IGC), instituição colaboradora do INSA neste estudo.

Nesta actualização do site foram inseridas **mais 34 sequências** obtidas no âmbito da solicitação do Ministério da Saúde para que fosse pesquisada a presença da nova variante recentemente identificada no Reino Unido (N501Y.V1; linhagem B.1.1.7) em amostras suspeitas, incluindo amostras associadas a casos positivos de COVID-19 com historial de viagem.

Entre as novas sequências (n=34), todas elas provenientes do arquipélago da Madeira, destaca-se a detecção de **18 casos associados à nova variante recentemente identificada no Reino Unido (N501Y.V1; linhagem B.1.1.7)* (Figura 1).**

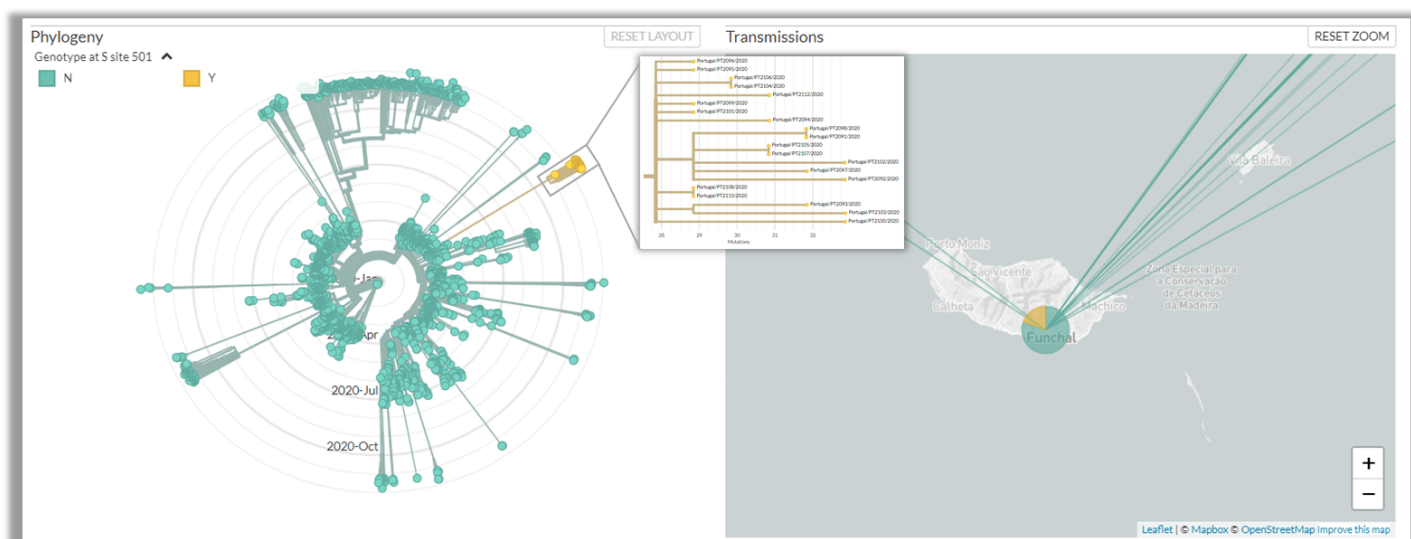


Figura 1. Enquadramento e dispersão geográfica da nova variante recentemente identificada no Reino Unido (N501Y.V1; linhagem B.1.1.7) no contexto dos genomas analisados em Portugal até à data. Os genomas com e sem a mutação N501Y associada a esta variante estão coloridos a amarelo e verde, respetivamente. **Até à data, foram analisadas 21 sequências desta variante (associadas a 18 casos de infecção).** *Para uma destas 21 amostras foi apenas obtida a sequência parcial do genoma, pelo que não foi integrada na árvore filogenética (<https://insaflu.insa.pt/covid19/>).



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



A Tabela seguinte e a Figura 2 apresentam a distribuição e frequência actualizadas das principais variantes genéticas do SARS-CoV-2 detectadas em Portugal desde o início de Novembro (n=437) (Tabela e Figura 3).

Variante	PROPORÇÃO (n=437) ^a	CONTEXTO
S: D614G+A222V (sub-clade "20A.EU1")	293 (67,0%)	- 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-infeção 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")	38 (8,7%)	- 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+N501Y+H69del/V70del++ (sub-clade 20B; N501Y.V1; linhagem B.1.1.7)	21 (4,8%) ^a	- Esta variante apresenta múltiplas mutações na proteína Spike, tendo sido detectada em Setembro no Reino Unido e, desde então, em múltiplos países. https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf - Esta variante apresenta uma deleção no gene S, a qual está associada à falha na detecção deste gene em alguns testes de RT-PCR. https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf
S: D614G+S98F (sub-clade 20A)	19 (4,3%)	- 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+N439Y+H69del/V70del (sub-clade 20A; linhagem B.1.258)	4 (0,9%) ^b	- A mutação N439Y na Spike altera o domínio de ligação ao receptor ("RBD") ACE2, podendo aumentar a afinidade da ligação da Spike às células do hospedeiro e alterar a reconhecimento por anticorpos. Weisblum et al, 2020, Elife. https://elifesciences.org/articles/61312; Barnes et al, 2020, Nature https://www.nature.com/articles/s41586-020-2852-1; Thomson et al, 2020, https://www.biorxiv.org/content/10.1101/2020.11.04.355842v1 - Esta variante apresenta uma deleção no gene S, a qual está associada à falha na detecção deste gene em alguns testes de RT-PCR. https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf
S: D614G+D80Y (sub-clade 20C)	4 (0,9%)	- 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
OUTRAS	58 (13,3%)	----

^a Para uma destas 21 amostras foi apenas obtida a sequência parcial do genoma, pelo que não foi integrada na árvore filogenética.

^b Para uma destas 4 sequências apenas foi possível obter parcialmente a sequência do gene S. Contudo, a partilha de outras mutações aponta que o posicionamento filogenético e classificação dessa sequência como "S: D614G+N439Y+H69del/V70del" é correcto.

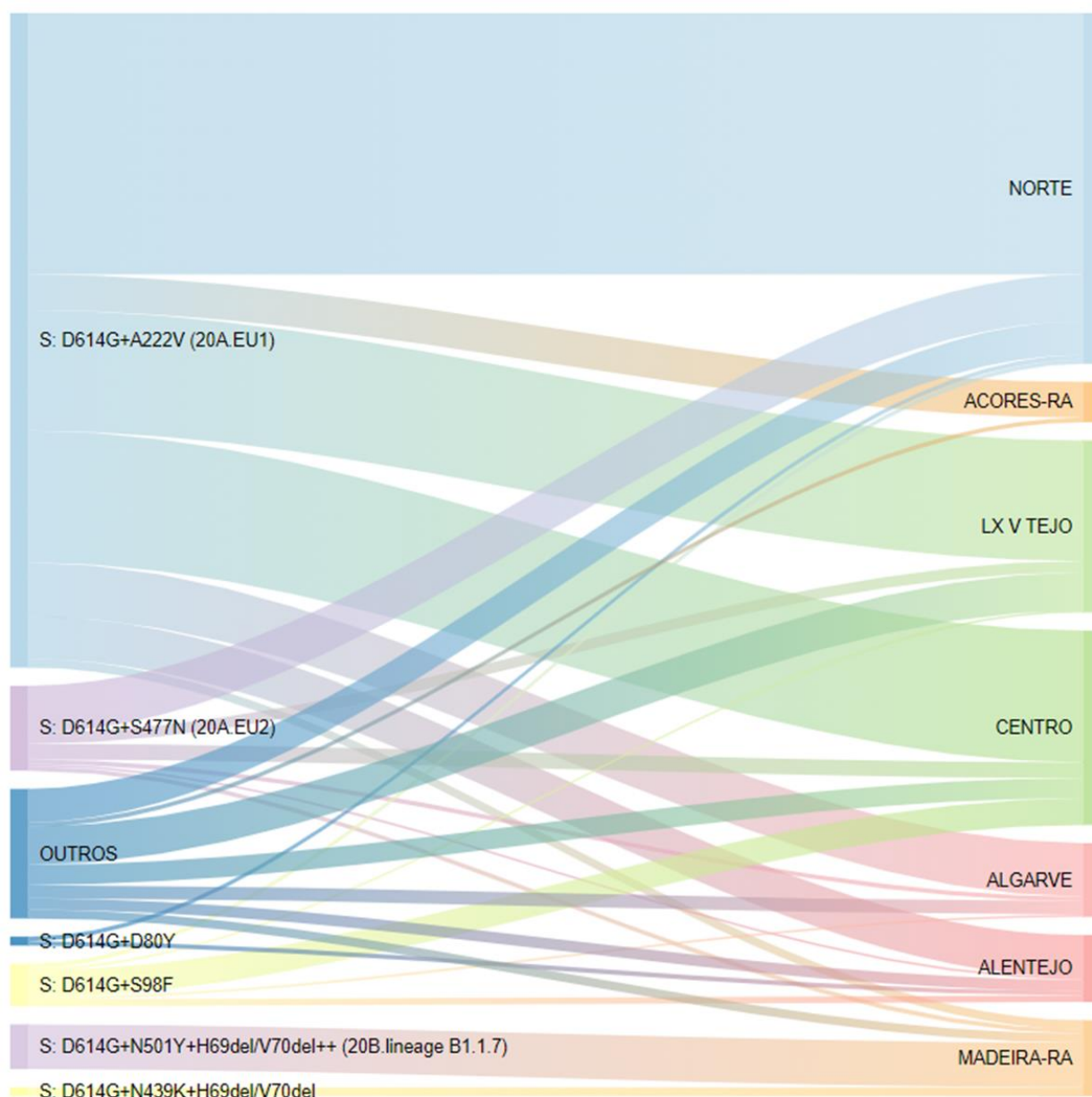


Figura 2. Frequência das principais variantes genéticas do SARS-CoV-2 detectadas em Portugal desde o início de Novembro (à esquerda) e sua distribuição por região (à direita).



OUTROS DESTAQUES

- Duas variantes agora detectadas, N501Y.V1 (linhagem B.1.1.7) e N439K (linhagem B.1.258), partilham uma deleção no gene S (levando à perda dos aminoácidos H69del/V70del da proteína Spike) (**Figura 3**), a qual está associada à falha na detecção deste gene em alguns testes de RT-PCR. Este artefacto não parece comprometer o diagnóstico laboratorial uma vez que esses testes se baseiam em vários alvos genéticos.

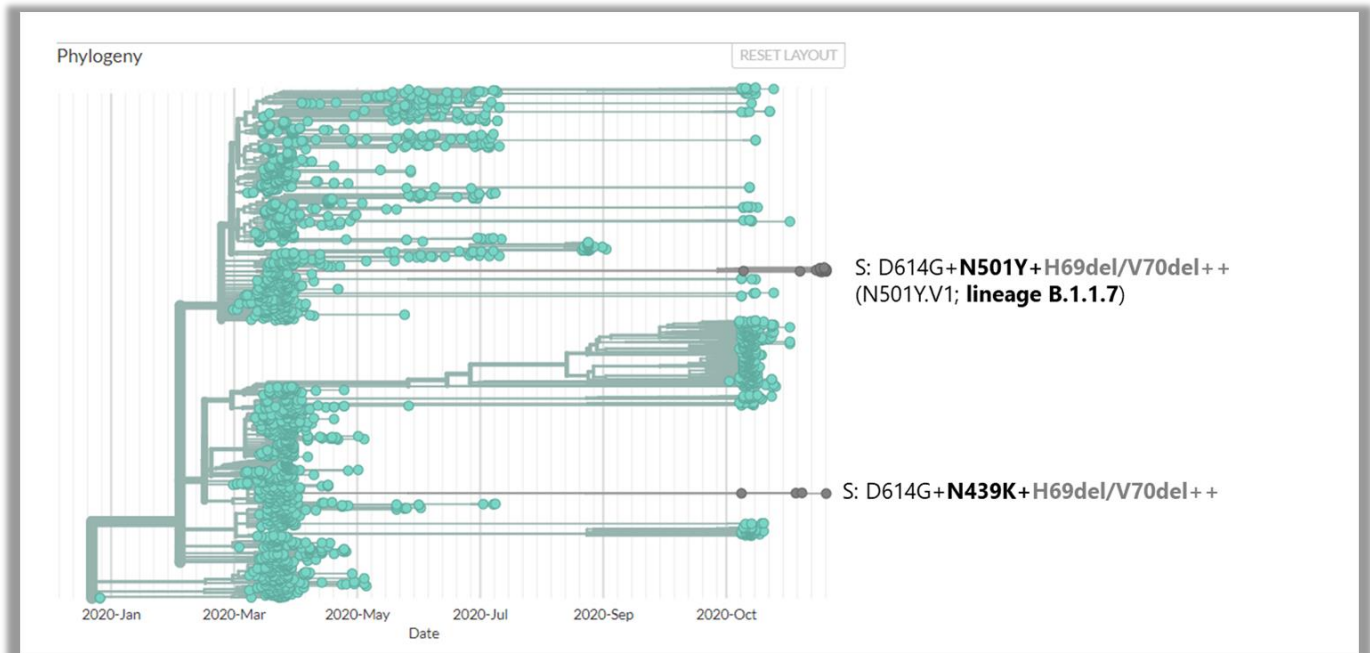


Figura 3. Enquadramento filogenético das sequências do genoma de SARS-CoV-2 com a deleção H69del/V70del na proteína Spike no contexto dos genomas analisados em Portugal até à data. Os genomas com e sem a deleção H69del/V70del estão coloridos a cinzento e verde, respetivamente.



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).