Instituto_Nacional de Saúde Doutor Ricardo Jorge
Observações_ Boletim Epidemiológico
2020 NÚMERO ESPECIAL 12
COVID-19
www.insa.pt
artigos breves_ n. 9
_Investigação da infeção por SARS-CoV-2

_Estudo da diversidade genética do SARS-CoV-2 (COVID-19) em Portugal

Study of SARS-CoV-2 (COVID-19) genetic diversity in Portugal

Vítor Borges[1], Joana Isidro[1], Helena Cortes-Martins[2], Sílvia Duarte[3], Luís Vieira[3], Raquel Guiomar[4], João Paulo Gomes[1]

j.paulo.gomes@insa.min-saude.pt

(1) Núcleo de Bioinformática. Departamento de Doenças Infeciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

(2) Unidade de Referência e Vigilância. Departamento de Doenças Infeciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

(3) Unidade de Tecnologia e Inovação. Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

(4) Laboratório Nacional de Referência para o Vírus da Gripe e outros Vírus Respiratórios. Departamento de Doenças Infeciosas, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

_Resumo

Os primeiros casos de COVID-19 em Portugal foram reportados no início de março, verificando-se rapidamente uma subida exponencial dos mesmos até ao início de abril. Neste âmbito, o Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA) iniciou o estudo da variabilidade genética do SARS-CoV-2 no país, criando um consórcio constituído por mais de 60 laboratórios, os quais possibilitaram a recolha de cerca de 3000 amostras positivas, de 159 concelhos. Nesta fase sequenciaram-se cerca de 1800 genomas de SARS-CoV-2, colocando Portugal como um dos países a nível mundial com maior empenho na caracterização das variantes genéticas do novo coronavírus em circulação. Os resultados obtidos possibilitaram a caracterização do início da epidemia em Portugal, permitindo, de alguma forma, avaliar a adequação de medidas de saúde pública que foram tomadas. O presente artigo debruça-se sobre os principais resultados obtidos neste âmbito, enumerando também uma série de estudos genéticos mais específicos que foram sendo realizados pelo INSA e com relevância em saúde pública.

_Abstract

The first cases of COVID-19 in Portugal were reported at the beginning of March, with an exponential increase rapidly occurring until the beginning of April. In this context, the Portuguese National Institute of Health (INSA) initiated the study of SARS-CoV-2 genetic variability in the country, creating a consortium of more than 60 laboratories, which made it possible to collect about 3000 positive samples from 159 counties. Currently, around 1800 SARS-CoV-2 genomes have been sequenced, placing Portugal as one of the countries worldwide with the greatest commitment to characterize the genetic variants of the new coronavirus in circulation. The results obtained enabled the characterization of the epidemic start in Portugal, and assess the adequacy of some public health measures. Although this article essentially focuses on results obtained pursuing this objective, it also lists a series of more specific genetic studies with Public Health relevance that have been carried out by INSA.

_Introdução

Em dezembro de 2019 foi reportado em Wuhan, China, o aparecimento de um elevado número de doentes com pneumonia viral de etiologia desconhecida. Foi mais tarde identificado um novo coronavírus, designado SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*), como responsável pela doença que se passou a designar por COVID-19 [(1-3)]. Rapidamente se alastrou a todas as regiões do globo e, em 11 de março de 2020, a Organização Mundial de Saúde (OMS) declarou o estado de pandemia. Segundo os dados da OMS (a 29 de novembro de 2020), foram já confirmados mundialmente mais de 61 milhões de casos, tendo originado cerca de 1,5 milhões de mortes [(4)]. Nesta mesma data, em Portugal, segundo os dados da Direção-Geral da Saúde (DGS), tinham sido já registados quase 300 mil casos e mais de 4400 mortes [(5)].

Após a sequenciação do primeiro genoma do SARS-CoV-2 de um doente de Wuhan [(2)] a comunidade científica empenhou-se fortemente no estudo da variabilidade genética deste vírus, sendo esta intenção igualmente expressa na Orientação 15/2020 da DGS publicada em março [(6)]. Neste sentido, o Departamento de Doenças Infeciosas do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA) lançou de imediato os alicerces para coordenar um estudo de âmbito nacional, com vista ao conhecimento das variantes genéticas a circular em Portugal. Embora tendo como objetivo principal a caracterização do início da epidemia no país, nomeadamente a identificação da cronologia e origem das introduções do SARS-CoV-2, prevendo aferir sobre o impacto das medidas de contenção, outros objetivos de carácter mais fundamental foram igualmente equacionados.

Entre eles, a possibilidade de determinar associações entre perfis mutacionais do SARS-CoV-2 e diferentes graus de severidade da COVID-19, bem como a determinação do grau de variabilidade genética de antigénios ou alvos de fármacos antivirais com possível impacto no desenvolvimento de medidas profiláticas (vacinas) e terapêuticas. O estudo teve aprovação pela Comissão de Ética para a Saúde do INSA e apoio financeiro por parte da Fundação para a Ciência e a Tecnologia e da Agência de Investigação Clínica e Inovação Biomédica.

_Objetivo

O presente artigo pretende ilustrar, de um modo resumido, a forma como está a decorrer o estudo (ainda em curso), os principais resultados já alcançados (também disponíveis em diversos relatórios divulgados publicamente) e as perspetivas para os próximos meses.

_Métodos

Obtenção de amostras positivas para COVID-19

De forma a otimizar o processo de recolha de amostras e facilitar a comunicação entre os intervenientes, foi criado um consórcio composto por cerca de 60 laboratórios, do sector público e privado, dispersos por Portugal continental e ilhas. Durante os primeiros meses da epidemia em Portugal, o INSA procedeu à recolha de cerca de 3000 amostras ou eluídos com resultados positivos para SARS-CoV-2 nestes laboratórios, e que foram armazenados, respetivamente, a -80ºC ou -20ºC até ao seu processamento.

Sequenciação total do genoma de SARS-CoV-2 e análise bioinformática

Os procedimentos subjacentes ao processamento das amostras/eluídos com vista à obtenção da sequência dos genomas virais ficaram a cargo do Núcleo de Bioinformática e da Unidade de Tecnologia e Inovação do INSA, sendo apoiados pela Unidade de Genómica do Instituto Gulbenkian de Ciência. Metodologicamente, de uma forma muito resumida, efetuaram-se procedimentos em *tandem*, desde a produção de cDNA a partir do RNA viral, PCR *multiplex* adaptado da *Artic Network* (*https://artic.network/ncov-2019*, *https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w*), preparação de bibliotecas genómicas com *kits* Illumina e, finalmente, a sequenciação de nova geração usando os equipamentos NextSeq (Illumina, Inc) e MiSeq (Illumina, Inc).

A análise genómica, da responsabilidade do Núcleo de Bioinformática, foi realizada essencialmente com recurso a uma *pipeline* bioinformática (INSaFLU, *https://insaflu.insa.pt/*) desenvolvida inicialmente por este grupo para a vigilância de base genómica da gripe sazonal e agora adaptada para manipulação de sequências do SARS-CoV-2. Acopladas a esta *pipeline* estão as plataformas bioinformáticas Nextstrain (*https://nextstrain.org/ncov*) e Microreact (*https://microreact.org/showcase*), as quais foram utilizadas para análise filogeográfica.

_Resultados e discussão

Balanço geral da variabilidade genética do SARS-CoV-2

Um dos primeiros passos que foi dado no sentido de garantir a rápida divulgação dos resultados, não só aos laboratórios pertencentes ao consórcio, mas também a toda a comunidade científica, foi a criação de um *website* (*https://insaflu.insa.pt/covid19/*) disponível publicamente, no qual são atualizados, em tempo real, todos os resultados que vão sendo obtidos.

Durante os primeiros meses da epidemia em Portugal foram analisadas pelo INSA cerca de 1800 sequências do genoma de SARS-CoV-2, representando todos os distritos do país e 159 concelhos, o que corresponde a 51,6% do total de concelhos nacionais. Em termos gerais, verificou-se que a distribuição por *clade* é semelhante àquela que é observada a nível europeu (*https://nextstrain.org/ncov/europe*) [7], refletida por uma esmagadora maioria dos vírus estudados (91,2%) a integrar o braço filogenético contendo os *clades* 20A (36,6%), 20B (52%) e 20C (2,6%). Estes *clades* apresentam, entre outros marcadores genéticos, a mutação “aa D614G” (“nt A23403G”), muito falada internacionalmente dada a sua rapidíssima disseminação mundial. Esta mutação afeta a

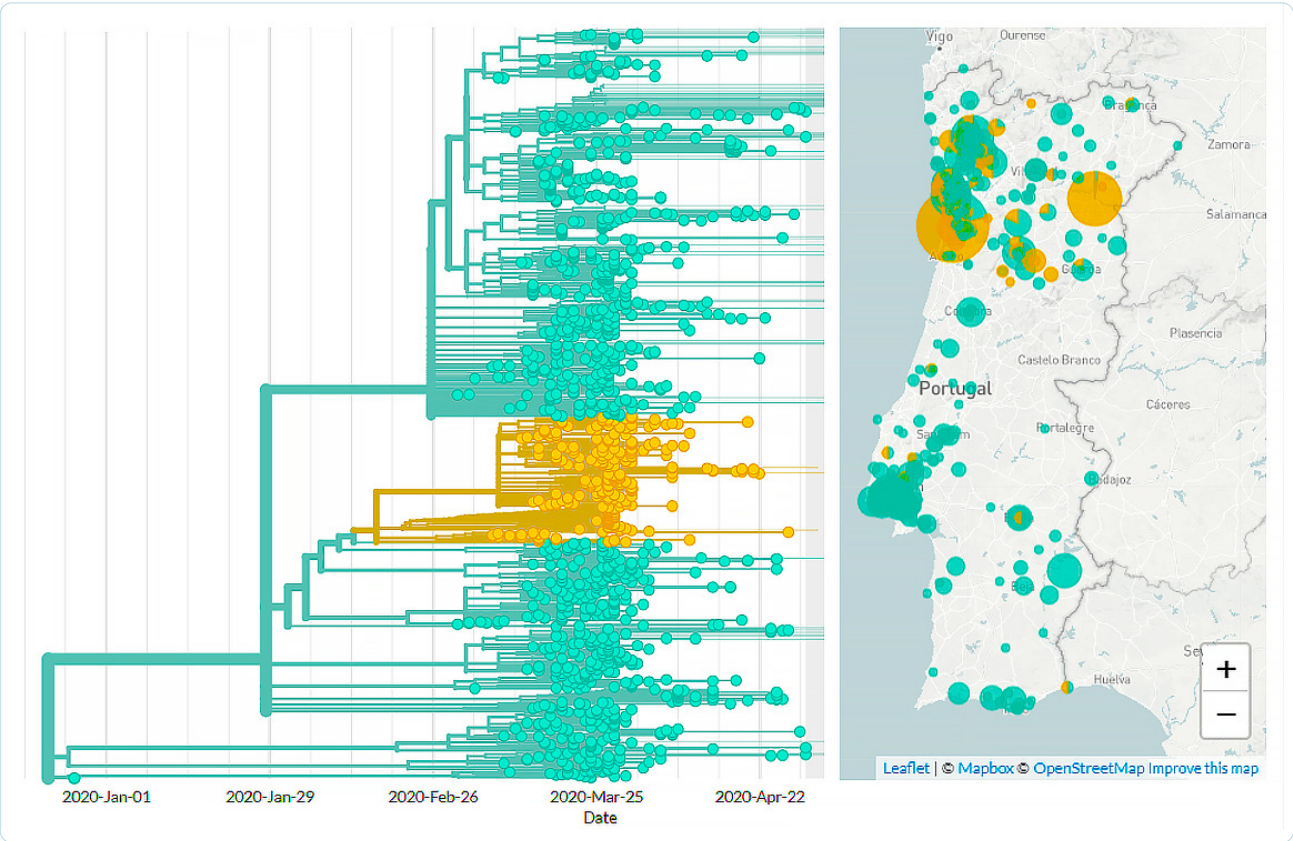
proteína "Spike (S)", responsável pela entrada do vírus SARS-CoV-2 nas células humanas, sendo também o principal antigénio deste vírus pandémico. De salientar que os vírus do *clade* 19B (2,4%) têm sido detetados maioritariamente em concelhos próximos da fronteira com Espanha (por exemplo, Arcos de Valdevez, 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.

Em termos de análise de mutações, verifica-se que os vírus a circular em Portugal apresentam uma taxa de mutação média de 2.3 mutações por genoma por mês, coincidente com a prevista para este vírus (*i.e.* 2 mutações por genoma por mês). É importante referir também que, no conjunto dos cerca de 1800 genomas virais analisados até à data, observou-se a ocorrência de 103 mutações distintas que alteram a proteína "Spike (S)", merecendo especial atenção dado o papel biológico desta proteína não só no processo de entrada do vírus na célula hospedeira, como também no desencadear da resposta imunitária por parte da pessoa infetada.

Caracterização do arranque da epidemia em Portugal

O resultado de maior destaque prende-se, sem dúvida, com a observação de que o início da pandemia em Portugal, com foco no Norte e Centro do país, foi caracterizado pelo espalhamento massivo de uma variante genética do SARS-CoV-2 com uma mutação (D839Y) na proteína Spike (figura 1)

Figura 1: Enquadramento e dispersão geográfica do sub-*clade* com a mutação D839Y na proteína Spike no contexto do global dos genomas analisados em Portugal até ao final de abril de 2020.



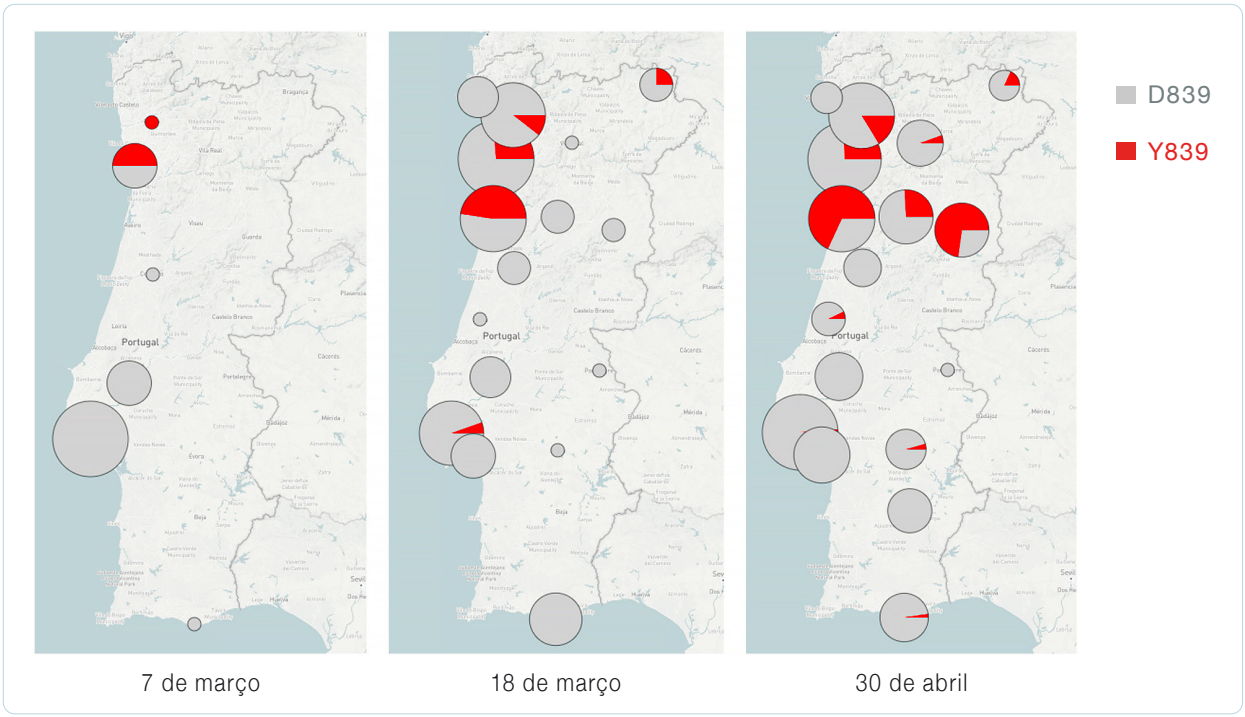
Os genomas com e sem a mutação D839Y estão coloridos, respetivamente, a amarelo e verde, sendo o tamanho dos círculos no mapa proporcional ao número de genomas sequenciados por localidade.

Esta mutação tinha sido já sinalizada internacionalmente como de particular interesse dado estar localizada na proteína Spike e, em particular, num domínio (*fusion peptide*) que se sabe ser crítico para a entrada do vírus nas células humanas. De salientar que esta mutação foi já detetada em pelo menos 13 países de quatro continentes. De acordo com os dados epidemiológicos fornecidos pelas autoridades de saúde pública e com os dados genéticos disponíveis publicamente na principal base de dados mundial para SARS-CoV-2 (GISAID), concluiu-se que esta variante genética do SARS-CoV-2 terá sido importada de Itália, região da Lombardia, pouco depois do meio de fevereiro. Após esta introdução, terá circulado em Portugal pelo menos uma semana antes de terem sido reportados os primeiros casos de COVID-19 no nosso país (2 de março), originando potencialmente cadeias de transmissão não detetadas.

Durante a fase exponencial da epidemia em Portugal, a variante contendo a mutação D839Y foi detetada em distritos do litoral das regiões Norte e Centro, tendo-se posteriormente disseminado para distritos no interior das mesmas regiões (figura 2). Esta variante foi responsável pelo grande surto em Ovar, tendo levado à cerca sanitária entre 17 de março e 17 de abril, a única a ser instituída em Portugal continental durante a primeira vaga. É razoável assumir que esta medida, juntamente com a proibição de circulação entre concelhos que teve lugar em alguns destes períodos, tenham constituído importantes medidas de saúde pública, contribuindo para a não disseminação desta variante genética para o Sul do país.

De salientar que a frequência relativa desta variante genética aumentou a uma taxa estimada de 12,1% (6,1%-18,2%, IC95%) a cada três dias, entre 14 de março e 9 de abril. Estima-se que, durante esse período, tenha sido responsável por cerca de 3800 casos (3177-4542 casos, IC95%) de COVID-19 em Portugal, correspondendo a cerca de um em cada quatro casos (24,8% (20,8-29,7%, IC95%) de COVID-19

Figura 2: Dispersão geográfica, por distrito, das frequências relativas da mutação D839Y (a vermelho) na proteína Spike, em três pontos temporais: 7 de março (deteção dos primeiros genomas com a mutação D839Y), 18 de março (quando foi declarado o estado de emergência) e 30 de abril de 2020.



As Regiões Autónomas dos Açores e da Madeira não são exibidas pois a variante D839Y não foi detetada nas mesmas.

Figura adaptada de: Borges V, *et al.* Emerg Microbes Infect. 2020 Dec;9(1):2488-96. https://doi.org/10.1080/22221751.2020.1844552

no país. Perante estes resultados, levantamos duas hipóteses que poderão, separada ou concomitantemente, ter contribuído para a elevada frequência da variante D839Y durante a fase exponencial da epidemia em Portugal. Por um lado, esta variante genética poderá ter tido maior oportunidade de propagação dado ter sido introduzida no país mais de uma semana antes dos primeiros casos detetados de COVID-19 (fenómeno designado cientificamente por *founder effect*). Por outro lado, o aumento significativo da sua frequência relativa poderá dever-se a um maior *fitness*, em particular em termos da sua capacidade de transmissão. No entanto, será necessária a realização de estudos funcionais que testem o impacto da mutação D839Y na capacidade infeciosa e de propagação do SARS-CoV-2.

Todos estes resultados e conclusões foram, entretanto, publicados numa revista internacional com *peer review* [8], num artigo do qual constam, como coautores, todos os colaboradores no âmbito do consórcio criado para este estudo.

Outros trabalhos desenvolvidos e perspetivas futuras

Em paralelo com esta linha de investigação principal, cujos resultados são o foco do presente artigo, outros trabalhos sobre a diversidade genética do SARS-CoV-2 foram desenvolvidos em paralelo no âmbito de novas colaborações entretanto estabelecidas. Assim, foram já realizados ou estão em curso os seguintes estudos:

i) Determinação do número de introduções do SARS-CoV-2 em Portugal e respetiva origem, que levaram ao estabelecimento da epidemia no nosso país (artigo em preparação);
ii) Participação em estudo europeu para avaliação da distribuição geográfica e temporal do SARS-CoV-2 (artigo publicado [7]);
iii) Sequenciação do genoma de SARS-CoV-2 para confirmação de um caso de transmissão vertical de SARS-CoV-2 (artigo publicado [9]);
iv) Sequenciação do genoma de SARS-CoV-2 no apoio à investigação de surtos nosocomiais (artigos em submissão);
v) Identificação das variantes genéticas de SARS-CoV-2 a circular em Portugal na segunda vaga da epidemia, em amostras positivas obtidas, durante o mês de novembro, na rede dos laboratórios colaboradores.

Finalmente, em termos de estudos a médio / longo prazo, o INSA procederá à caracterização dos perfis mutacionais do SARS-CoV-2 que sejam identificados em pessoas que, apesar de terem sido vacinadas, tenham contraído a infeção ("falências vacinais"). A maior ou menor frequência de determinadas variantes genéticas associadas a falências vacinais poderá determinar a eventual necessidade de adaptações / otimizações das vacinas existentes.

Dos resultados obtidos até ao momento no âmbito deste estudo, é possível inferir a mais-valia dos estudos de epidemiologia molecular no conhecimento da disseminação dos novos agentes infeciosos e respetiva dinâmica. Ainda, destaca-se a elevada capacidade nacional, concretamente do INSA, para a sequenciação e análise genómica de microrganismos usando tecnologia de ponta e o espírito colaborativo dos laboratórios nacionais e das Autoridades de Saúde, no desígnio de contribuir para um conhecimento mais profundo desta epidemia.

Agradecimentos:

Os autores deste artigo dirigem um especial agradecimento ao consórcio *Portuguese network for SARS-CoV-2 genomics*, o qual engloba todos os laboratórios a nível nacional que enviaram amostras para caracterização genética do SARS-CoV-2, a equipa do INSA envolvida no diagnóstico laboratorial da COVID-19, a equipa da Unidade de Genómica do Instituto de Gulbenkian Ciência e colaboradores no campo da bioinformática (Hugo Martiniano e Miguel Pinheiro). Agradece-se, ainda, de forma particular, às autoridades de saúde nacionais, regionais e locais pela partilha de dados demográficos e epidemiológicos.

Financiamento:

Este estudo é cofinanciado pela Fundação para a Ciência e Tecnologia e pela Agência de Investigação Clínica e Inovação Biomédica (234_596874175) no âmbito da *call* RESEARCH 4 COVID-19.

Parte dos resultados apresentados neste artigo correspondem a uma versão reduzida em português do artigo: Borges V, Isidro J, Cortes-Martins H, Duarte S, Vieira L, Leite R, Gordo I, Caetano CP, Nunes B, Sá R, Oliveira A, Guiomar R; Portuguese network for SARS-CoV-2 genomics, Gomes JP. Massive dissemination of a SARS-CoV-2 Spike Y839 variant in Portugal. Emerg Microbes Infect. 2020 Dec;9(1):2488-96. https://doi.org/10.1080/22221751.2020.1844552

Referências bibliográficas:

(1) Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270-273. https://doi.org/10.1038/s41586-020-2012-7.

(2) Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020 Mar;579(7798):265-269. https://doi.org/10.1038/s41586-020-2008-3. Erratum in: Nature. 2020 Apr;580 (7803):E7.

(3) Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020 May;20(5):533-534. https://doi.org/10.1016/s1473-3099(20)30120-1. Erratum in: Lancet Infect Dis. 2020 Sep;20(9):e215.

(4) World Health Organization. Coronavirus disease (COVID-19) pandemic [Internet]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

(5) Direção-Geral da Saúde. COVID-19 [Internet]. https://covid19.min-saude.pt/

(6) Direção-Geral da Saúde. Orientação nº 015/2020, de 23/03/2020 (atualizada a 24/04/2020). COVID-19: Diagnóstico Laboratorial - Diagnóstico laboratorial; produtos biológicos; SARS-CoV-2; COVID-19. https://www.dgs.pt/directrizes-da-dgs/orientacoes-e-circulares-informativas/orientacao-n-0152020-de-23032020.aspx

(7) Alm E, Broberg EK, Connor T, et al.; WHO European Region sequencing laboratories and GISAID EpiCoV group; WHO European Region sequencing laboratories and GISAID EpiCoV group. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill. 2020 Aug;25(32):2001410. https://doi.org/10.2807/1560-7917.es.2020.25.32.2001410. Erratum in: Euro Surveill. 2020 Aug;25(33):200820c.

(8) Borges V, Isidro J, Cortes-Martins H, et al; Portuguese network for SARS-CoV-2 genomics, Gomes JP. Massive dissemination of a SARS-CoV-2 Spike Y839 variant in Portugal. Emerg Microbes Infect. 2020 Dec;9(1):2488-2496. https://doi.org/10.1080/22221751.2020.1844552

(9) Correia CR, Marçal M, Vieira F, et al. Congenital SARS-CoV-2 Infection in a Neonate With Severe Acute Respiratory Syndrome. Pediatr Infect Dis J. 2020 Dec;39(12):e439-e443. https://doi.org/10.1097/inf.0000000000002941