

## SARS-CoV-2 introductions and early dynamics of the epidemic in Portugal

Vítor Borges<sup>1,50</sup>, Joana Isidro<sup>1,50</sup>, Nídia Sequeira Trovão<sup>2,50</sup>, Sílvia Duarte<sup>3</sup>, Helena Cortes-Martins<sup>4</sup>, Hugo Martiniano<sup>3</sup>, Isabel Gordo<sup>5</sup>, Ricardo Leite<sup>5</sup>, Luís Vieira<sup>3,6</sup>, Portuguese network for SARS-CoV-2 genomics (Consortium)\*, Raquel Guiomar<sup>7</sup> & João Paulo Gomes<sup>1</sup>✉

### Abstract

**Background** Genomic surveillance of SARS-CoV-2 in Portugal was rapidly implemented by the National Institute of Health in the early stages of the COVID-19 epidemic, in collaboration with more than 50 laboratories distributed nationwide.

**Methods** By applying recent phylodynamic models that allow integration of individual-based travel history, we reconstructed and characterized the spatio-temporal dynamics of SARS-CoV-2 introductions and early dissemination in Portugal.

**Results** We detected at least 277 independent SARS-CoV-2 introductions, mostly from European countries (namely the United Kingdom, Spain, France, Italy, and Switzerland), which were consistent with the countries with the highest connectivity with Portugal. Although most introductions were estimated to have occurred during early March 2020, it is likely that SARS-CoV-2 was silently circulating in Portugal throughout February, before the first cases were confirmed.

**Conclusions** Here we conclude that the earlier implementation of measures could have minimized the number of introductions and subsequent virus expansion in Portugal. This study lays the foundation for genomic epidemiology of SARS-CoV-2 in Portugal, and highlights the need for systematic and geographically-representative genomic surveillance.

### Plain language summary

Analysing SARS-CoV-2 genetic material and how it changes over time can help us understand how the virus spreads between countries and determine the impact of control measures. In this study, we investigated SARS-CoV-2 transmission and evolution in the early stages of the COVID-19 pandemic in Portugal. In particular, we reconstructed the routes and timeliness of viral introductions into the country and assessed the relative contribution of each introduction in terms of how the epidemic evolved over time. We detected at least 277 independent introductions, mostly from European countries (namely the United Kingdom, Spain, France, Italy, and Switzerland), which were consistent with the countries with the highest connectivity with Portugal. This study reflects an unprecedented effort in the field of the infectious diseases in Portugal, highlighting the need for systematic and geographically-representative surveillance to aid public health efforts to control the virus.

<sup>1</sup>Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal. <sup>2</sup>Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. <sup>3</sup>Innovation and Technology Unit, Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal. <sup>4</sup>Reference and Surveillance Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal. <sup>5</sup>Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal. <sup>6</sup>Centre for Toxicogenomics and Human Health (ToxOmics), Genetics, Oncology and Human Toxicology, Nova Medical School; Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisbon, Portugal. <sup>7</sup>National Reference Laboratory for Influenza and other Respiratory Viruses, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal. <sup>50</sup>These authors contributed equally: Vítor Borges, Joana Isidro, Nídia Sequeira Trovão. \*A list of authors and their affiliations appears at the end of the paper. ✉email: [j.paulo.gomes@insa.min-saude.pt](mailto:j.paulo.gomes@insa.min-saude.pt)

**S**ARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), the causative agent of COVID-19, is a novel betacoronavirus that was first reported in December 2019 in Wuhan, China<sup>1,2</sup>. By 29 March 2021, it had already caused more than 126 million cases and 2.7 million deaths worldwide<sup>3,4</sup>. In order to control the virus arrival and spread, many countries adopted rigid public health measures, including complete border closures and general lockdowns, with tremendous consequences at economic and social levels. At the early stages of an epidemic, the success of public health measures is particularly dependent on their timely implementation, which requires comprehensive diagnosis/surveillance systems that are able to efficiently trace where the virus is being introduced and circulating<sup>5-7</sup>. Taking advantage of the extraordinary advances in sequencing technologies, modern surveillance systems are progressively relying on genomic epidemiology as a crucial tool for outbreak investigation and for tracking virus evolution and spread<sup>7-9</sup>. Genomic surveillance of SARS-CoV-2 can be particularly useful to: (i) understand the contribution of “new introductions” versus “local transmission” to the number of new cases at continent/country/regional levels; (ii) evaluate the impact of non-pharmaceutical interventions on the outcomes of transmission chains; (iii) characterize the genetic variability that may negatively affect molecular diagnostic tests; (iv) monitor genetic variability affecting antigens and targets of antiviral drugs with potential impact on the development/effectiveness of prophylactic (vaccines) and therapeutic measures; and (v) investigate potential associations between genetic variants and infectious load, patient immunological status, clinical outcomes (e.g., infection duration, disease severity, etc.)<sup>5,10</sup>.

Acting as the National Reference Laboratory for SARS-CoV-2, the Portuguese National Institute of Health (INSA) Doutor Ricardo Jorge rapidly established a genome-based molecular surveillance strategy for SARS-CoV-2 in Portugal, setting up a large nationwide network involving more than 50 laboratories. A bilingual website (<https://insaflu.insa.pt/covid19>) was launched, providing updated data regarding the analysis of the SARS-CoV-2 genetic diversity and geotemporal dynamics, based on state-of-the-art methodologies for real-time tracking pathogen evolution<sup>11,12</sup>. Also, “situation reports” with major highlights are being released periodically to participating laboratories, national and regional public health authorities, and other stakeholders.

Despite all the advantages of genomic surveillance, the uneven geographic sampling of viral genomes can severely skew phylogeographic inferences based on discrete trait ancestral reconstruction<sup>13</sup>, therefore hindering the ability to accurately trace the seeding and dissemination patterns of SARS-CoV-2. The COVID-19 pandemic has been characterized by an unprecedented amount of genomic data and associated metadata, such as information on the patients’ recent movements prior to having developed any symptoms.

In the present study, we reconstruct and characterize the spatio-temporal dynamics of SARS-CoV-2 introductions and early dissemination in Portugal using newly developed phylogenetic models that allow integration of individual-based travel history, in order to obtain a more realistic reconstruction of the viral dynamics<sup>13</sup>. This includes inferences of the timelines of the first introductions, geographic location of ancestral lineages, and the contribution of detected introductions to the epidemic evolution.

## Methods

**Sample characterization.** Samples used in this study were collected as part of the ongoing national SARS-CoV-2 laboratory surveillance conducted by INSA, Portugal, in collaboration with

Instituto Gulbenkian de Ciência (IGC). SARS-CoV-2 positive samples (either clinical specimens or RNA) were provided by a nationwide network, consisting of more than 50 laboratories, that was established at the beginning of the epidemic in Portugal. Anonymized date of sample collection, date of illness onset, and travel history were provided by laboratories and Regional and National Health Authorities. Geographical data presented in this study refers to the Region (“Health Administration region”) of the patients’ residence or, when no information was available, to the Region of exposure or of the hospital/laboratory that collected/sent the sample.

**SARS-CoV-2 genome sequencing and assembly.** SARS-CoV-2 positive RNA samples were subjected to genome sequencing using a whole-genome amplification strategy with tiled, multiplexed primers<sup>14</sup> and the ARTIC Consortium protocol (<https://artic.network/ncov-2019>; <https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmui6w>), with slight modifications, as previously described<sup>15</sup>. Analysis of sequence read data was conducted using the bioinformatics pipeline implemented in INSAFLU (<https://insaflu.insa.pt/>; <https://github.com/INSAFLU>), which is a web-based (and also locally installable) platform for amplicon-based next-generation sequencing data analysis<sup>16</sup>. Sequence inspection and validation was performed as previously described<sup>15</sup>.

**Classification by clades and lineages.** We explored the diversity of INSA sequences using a variety of nomenclature strategies, namely Nextstrain (using <https://clades.nextstrain.org/>; 9 November 2020), GISAID (<https://www.gisaid.org/>; 23 July 2020) and Phylogenetic Assignment of Named Global Outbreak LINEages (cov-lineages.org) (<https://pangolin.cog-uk.io/>; 16 October 2020)<sup>17</sup>. While Nextstrain and GISAID clade nomenclatures provide a less detailed categorisation of globally circulating diversity, cov-lineages.org classification is focused on identifying highly specific lineages that are actively transmitting in the population. Classification is provided in Supplementary material (Supplementary Data 1).

**Assessment of genome sequencing by country.** To assess the contribution of each country to the set of publicly available SARS-CoV-2 genomes and to determine the proportion of the number of genomes on the total number of reported COVID-19 cases (genome sampling) of a given country during the study period (until 31 March 2020), we obtained the number of cases per country from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University ([https://github.com/CSSEGISandData/COVID-19/blob/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/time\\_series\\_covid19\\_confirmed\\_global.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv)) and the number of genomes from GISAID (by 8 August 2020). Only the genomes with collection date until 31 March 2020 were considered. When a given genome lacked the collection day date, only specifying the month of collection, it was assigned to the last day of the respective month. When the number of genomes on a given day was higher than the number of cases, the number of cases was considered for graphical representation. Final data on the assessment of genome sampling by country is provided in Supplementary material (Supplementary Data 2).

**Selecting a genomic background dataset.** For the phylogenetic analyses, we downloaded full-length viral genome sequences from GISAID (<https://www.gisaid.org/>) on 6 August 2020 with collection dates before 1 April 2020 (Supplementary Data 2). For computational efficiency in the downstream operations, we

analysed the A and B lineages separately. Multiple sequence alignments with a reference genome (MN908947.3) were performed using MAFFT v7.458 with parameter `--addfragments`<sup>18</sup>. Sequences with fewer than 75% unambiguous bases were excluded, as well as duplicate sequences defined as having identical nucleotide composition, collected on the same date and in the same country. The resulting dataset was trimmed at the 5' and 3' ends resulting in a multi-sequence alignment with 29,780 nucleotides. Sequences with date information only at the year-level were also excluded. This dataset was subjected to multiple iterations of phylogeny reconstruction using IQ-TREE multicore software version 1.6.12<sup>19</sup> with parameters `-m GTR+G`, and exclusion of outlier sequences whose genetic divergence was incongruent with sampling date using TempEst software version 1.5.2<sup>20</sup>, resulting in 1632 and 22,124 sequences for the A and B datasets, respectively. GISAID acknowledgment table for the background dataset is provided as Supplementary Material (Supplementary Data 3).

**Subsampling strategy.** The magnitude of the lineage B datasets prohibits a full Bayesian inference approach in a reasonable timeframe. To overcome this constraint, we used a subsampling strategy that removes sequences such that monophyletic clusters that consist entirely of sequences from a particular country are represented by a single sequence. The excess sequences in a country-specific monophyletic clade do not contribute any additional information to the between-country diffusion process we aim to infer<sup>21</sup>. This process resulted in a dataset with 13,489 sequences (B\_CS). Despite the almost 40% downsampling in the B lineage dataset, its size is still excessive for timely computational inferences. To further address this issue, we have built a phylogeny using IQ-TREE, as described previously, and partitioned the tree into six monophyletic clades (B\_CS1 through B\_CS6). These clades were examined for outlier sequences whose genetic divergence and sampling date were incongruent using TempEst software version 1.5.2<sup>20</sup>. All data sets exhibited a positive correlation between genetic divergence and sampling time and appear to be suitable for phylogenetic molecular clock analysis<sup>20</sup> (Supplementary Fig. 1).

**Bayesian evolutionary inference of SARS-CoV-2 detected in Portugal.** A total of 1275 SARS-CoV-2 genome sequences (obtained from positive samples collected until 31 March 2020) from Portugal were analysed in this study (INSA's collection, as of 23 July 2020; Supplementary Data 1). Our interest lies in estimating the viral evolutionary history and spatial diffusion process during the early epidemics in the country. Travel history data is of particular importance when analyzing low diversity data, such as that for SARS-CoV-2, using Bayesian joint inference of sequence and location traits because sharing the same location state can contribute to the phylogenetic clustering of taxa<sup>13</sup>. For each of the datasets (A, and B\_CS1 through B\_CS6), we performed a joint genealogical and phylogeographic inference of time-measured trees using Markov chain Monte Carlo (MCMC) sampling implemented in the Bayesian Evolutionary Analysis Sampling Trees (BEAST) package<sup>22</sup>. We applied a Hasegawa-Kishino-Yano 85 (HKY85)<sup>23</sup> substitution model with gamma-distributed rate variation among sites<sup>24</sup>. We used an uncorrelated lognormal relaxed molecular clock to account for evolutionary rate variation among lineages<sup>25</sup> and specified an exponential growth coalescent prior in our analyses.

To integrate the travel history information obtained from (returning) travelers, we followed Lemey et al.<sup>13</sup> and augmented the phylogeny with ancestral nodes that are associated with a location state (but not with a known sequence), and enforced the

ancestral location at a point in the past of a lineage. We specified normal prior distributions on the travel times informed by an estimate of time of infection and truncated to be positive (back-in-time) relative to sampling date. Specifically, we used a period of 14 days (incubation period of 99% of patients<sup>26</sup> where travel history information was collected for all recent movements), and a period between symptom onset and testing with an estimated mean of 4.70 days for the patients in the INSA cohort (estimated from data available for 717/1275 individuals), and a standard deviation of 4.06 days to incorporate the uncertainty on the period between symptom onset and testing. The location traits associated with taxa and with the ancestral nodes were modeled using a bidirectional asymmetric discrete diffusion process<sup>27</sup>. We ran and combined at least eight independent MCMC analyses for 50 million generations, sampling every 50,000th generation and removed 10% as chain burn-in. Stationarity and mixing were investigated using Tracer software version 1.7.1<sup>28</sup>, making sure that effective sample sizes for the continuous parameters were greater than 100. We used the high-performance computational capabilities of the Biowulf cluster at the National Institutes of Health (Bethesda, MD, USA) (<http://biowulf.nih.gov>) to perform these analyses. Portuguese clusters were assumed for phylogeographic summaries if their topology posterior probability was  $\geq 0.001$ . If the excluded genomes had known travel history, they were re-integrated along with same-cluster sequences if those did not cluster in a clade-defining polytomy (recovered a total of 14 sequences, 7 of them with travel history). The location of the most recent common ancestor (MRCA) of Portuguese clusters is usually inferred as Portugal, thus we compared the estimated locations and times for the parent nodes of the MRCA (for simplicity, here on referred to as PMRCA) across Portuguese BEAST clades representing the origin and timing of seeding events into Portugal.

**Real-time data sharing of SARS-CoV-2 genetic diversity and geotemporal spread in Portugal.** A website (<https://insaflu.insa.pt/covid19>) was launched on 28 March 2020 for real-time data sharing on SARS-CoV-2 genetic diversity and geotemporal spread in Portugal. This site gives access to “situation reports of the study and provides interactive data navigation using both Nextstrain (<https://nextstrain.org/>)<sup>11</sup> and Microreact (<https://microreact.org/>)<sup>12</sup> tools. As of 23 July 2020, genomic and phylogenetic analysis were performed using the SARS-CoV-2 Nextstrain pipeline version from 23 March 2020 (<https://github.com/nextstrain/ncov>), with slight modifications<sup>15</sup>. For data navigation, an IQ-TREE-derived<sup>19</sup> phylogenetic tree enrolling the 1275 studied sequences, and the associated metadata, can be visualized interactively at <https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>.

**Ethical approval.** Samples were obtained in the frame of the ongoing national SARS-CoV-2 genomic surveillance coordinated by the Portuguese National Institute of Health (INSA), being collected as part of the routine clinical care and laboratory procedures of the laboratories/hospitals (“Portuguese network for SARS-CoV-2 genomics”) collaborating in this system. This study was approved by the Ethical Committee (“Comissão de Ética para a Saúde”) of INSA, dismissing the need for individuals’ informed consent. Designations of all genome sequences are fully anonymized, and no identifying information of the associated patients is provided. Anonymized date of sample collection, date of illness onset and travel history were provided by laboratories and Regional and National Health Authorities.



**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Results

**Epidemiological trends and circulating diversity during the early COVID-19 pandemic in Portugal.** The first COVID-19 confirmed cases in Portugal were reported on 2 March 2020 after laboratory confirmation by the National Institute of Health (INSA) Doutor Ricardo Jorge. As the COVID-19 epidemic progressed in the country, INSA, as the National Reference Laboratory, gradually supported and validated the extension of the laboratory network throughout the country, where 37 laboratories were already set up to perform molecular testing of SARS-CoV-2 by March 31, 2020. With COVID-19 cases exponentially increasing in Portugal (Fig. 1a) and Europe (with alarming trends in Italy and Spain)<sup>4</sup>, Portugal adopted rigid public health measures, including suspension of flights from/to Italy (10 March 2020), closure of land borders and schools (16 March 2020), suspension of flights to non-EU countries and general lockdown (18 March 2020). INSA rapidly implemented and coordinated the nationwide genomic surveillance of SARS-CoV-2 (<https://insaflu.insa.pt/covid19>) with a particular focus on providing a comprehensive picture of the introductions, genetic diversification, and propagation of SARS-CoV-2 during the early-stage pandemic at a country scale. A total of 1275 SARS-CoV-2 positive samples collected in March were successfully subjected to virus genome sequencing (Supplementary Data 1), which corresponds to 15.5% (1275/8251) of all COVID-19 cases confirmed during the first month of the pandemic in Portugal (Fig. 1b). The viral genomic sequence sampling ranged from 10.7% (523/4910) in the Northern Health Administration region (the region with the highest number of confirmed cases) to 85.4% (41/48) in Madeira Archipelago (the region with the lowest number of confirmed cases) (Fig. 1c). Despite the delayed epidemic trajectory in comparison with other European countries, Portugal was among the countries generating the highest volumes of SARS-CoV-2 genomic data during the early COVID-19 pandemic (Fig. 1d and Supplementary Data 2).

The relative frequency trends of SARS-CoV-2 clades and lineages in Portugal during March 2020 (Supplementary Fig. 2) did not differ significantly from the scenario observed at the European level<sup>5</sup>. Most of the analyzed genomes (89.8%) belong to the 20A/G, 20B/GR, and 20C/GH clades (Nextstrain/GISAID nomenclatures), which display, among other genetic markers, the D614G amino acid replacement in the Spike protein. The SARS-CoV-2 G614 variant was most likely introduced in Europe by mid-January/early February 2020 and became dominant at a worldwide level<sup>29</sup>. Clades 19A/L/V/O (dominant during early pandemic in China) and 19B/S (rare in Europe, with the notable exception of Spain)<sup>5</sup> were found at the relative frequencies around 7.5% and 2.8% in March, respectively, and showed a decreasing frequency trend similarly to what has been observed at the global level. When applying the Pango lineages ([cov-lineages.org](https://covid19.hki.se/nomenclature/index.html)) nomenclature, the main B lineage (roughly covering 19A/20A/20B/20C Nextstrain clades) was dominant in March, showing ample diversification into sublineages. Among these, it is highlighted the increased frequency of the B.1.1 sublineages (roughly corresponding to clade 20B/GR), as well as of the B.1, B.1.5, and B.1.91 lineages (all mostly including 20A/G virus). Of note, the B.1.91 sub-lineage, and part of its ancestor sublineage B.1, correspond to the Spike G614 + Y839 variant that massively disseminated in Portugal during the early epidemic (22% and 59% of the sampled genomes from the North and Center regions of Portugal by 30th April)<sup>15</sup>.

**Introductions and early spread of SARS-CoV-2 in Portugal.** In order to assess the origins and measure the number of SARS-CoV-2 introductions in Portugal prior to 31 March 2020, we performed Bayesian phylodynamic analyses integrating the rich travel history data. The global TMRCA for the A lineage tree ranged between 11 November 2019–18 December 2019 (95% highest posterior density) and the global TMRCA for the B lineage trees ranged between 11 November 2019–3 February 2020 (95% highest posterior density). Travel history could be collected for 619 (48.5%) out of 1275 confirmed cases detected in March, with 128 (10.0%) reporting international travels within the potential incubation period (i.e., 14 days before clinical onset date) and 491 (38.5%) reporting no travels within the same period (Fig. 2a).

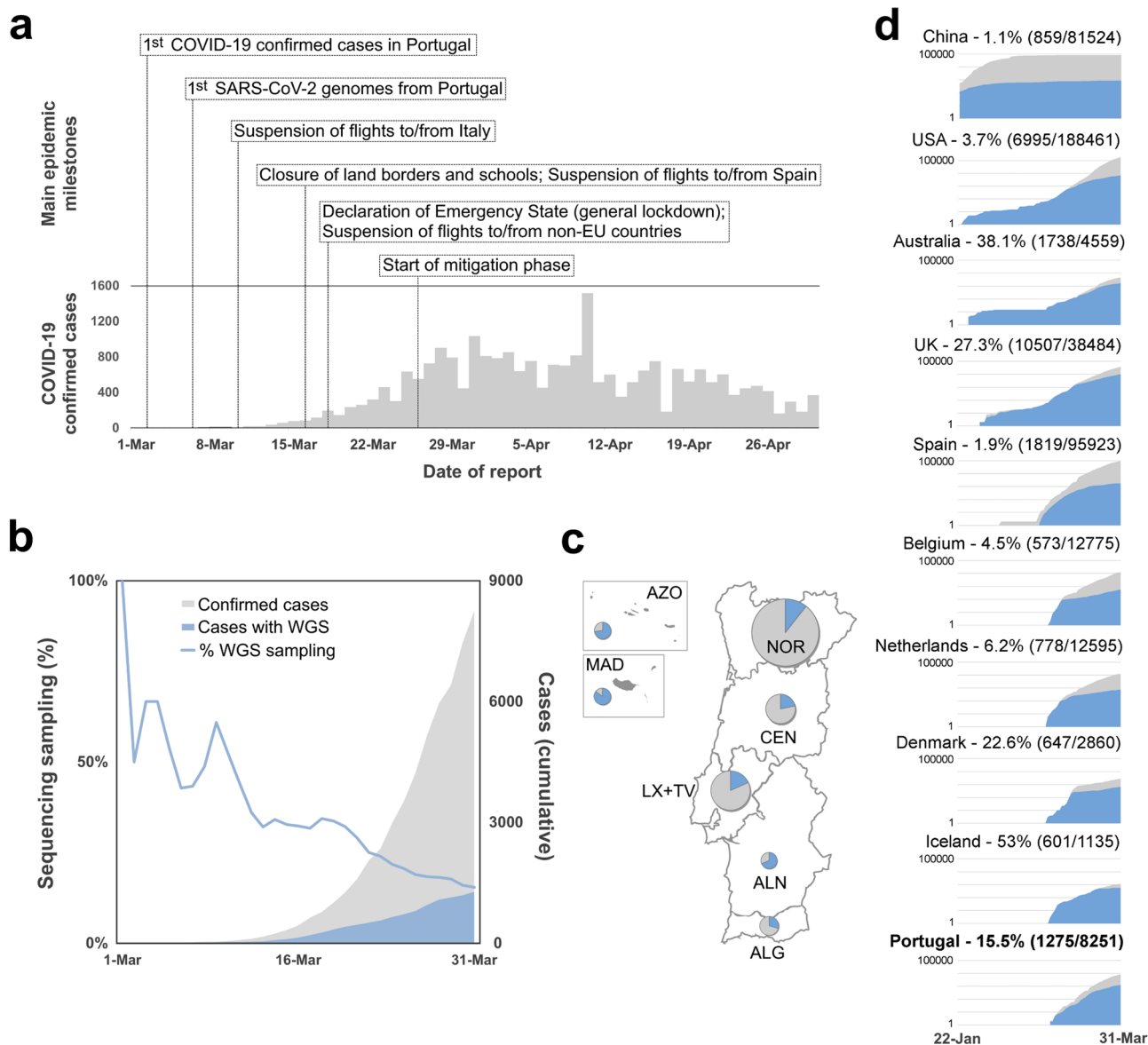
Overall, we detected 277 clades (16 from lineage A and 261 from lineage B) representing SARS-CoV-2 introductions in Portugal until 31 March 2020, involving a total of 979 out of the 1275 sequences analyzed. Of these, 296/979 genomes belong to clades that include sequences from cases with known travel history (Fig. 3b). Overall, the phylogeographic reconstruction revealed that UK, France, and Spain seeded 69% of all introductions into Portugal (Fig. 3), mostly to the Lisbon and Tagus Valley region, which was estimated to have received ~30% of all introductions, followed by the North (27%), Center (10%), Algarve (4%), Azores (2%), Alentejo (2%), and Madeira (1%) regions (Figs. 4 and 5A and Supplementary Data 4).

We also examined the number of infections generated by introductions from different countries (Figs. 3, 4, and 5B and Supplementary Data 5). Interestingly, despite the UK, France, and Spain having seeded the majority of introductions, at least one seeding event linked to travel history from Italy generated the largest outbreak recorded in the early stages of the COVID-19 epidemic in Portugal (cfr. below)<sup>15</sup>. We noticed a similar pattern for an introduction from Thailand, however, this event had a low supported topology posterior probability (Figs. 4 and 5B and Supplementary Data 5).

The spatial analysis reconstructed 16 introductions of lineage A into Portugal encompassing 33 Portuguese sequences. The majority of this lineage viral flow was seeded by Spain (69%), with 41% of all of these being introduced into Lisbon & Tagus Valley, 12% into the Center and Alentejo regions, respectively, and 6% into the North region. The remainder of all lineage A seeding events were estimated to have originated from Australia, France, Italy, Oman, Saudi Arabia and were mostly captured by the ability to integrate travel history data into the phylogeographic reconstruction (Supplementary Fig. 3 and Supplementary Data 6). The introductions resulted in transmission clusters of varying size, with the largest, seeded by Spain, generating a transmission chain of at least 7 cases in Alentejo.

A similar phylogeographic analysis for lineage B resulted in the reconstruction of 261 introductions covering 946 Portuguese genomes. The majority of lineage B introductions was seeded by the United Kingdom (UK) (36%), Spain (20%), and France (14%). The introductions resulted in transmission clusters of varying sizes. The largest cluster was seeded by Italy and generated a transmission chain of at least 252 cases in the North and Center regions. Switzerland (4%), United Arab Emirates (3%), United States of America (3%), Netherlands (2%) also contributed to the remainder of lineage B introductions and establishment in the country (Supplementary Fig. 4 and Supplementary Data 7).

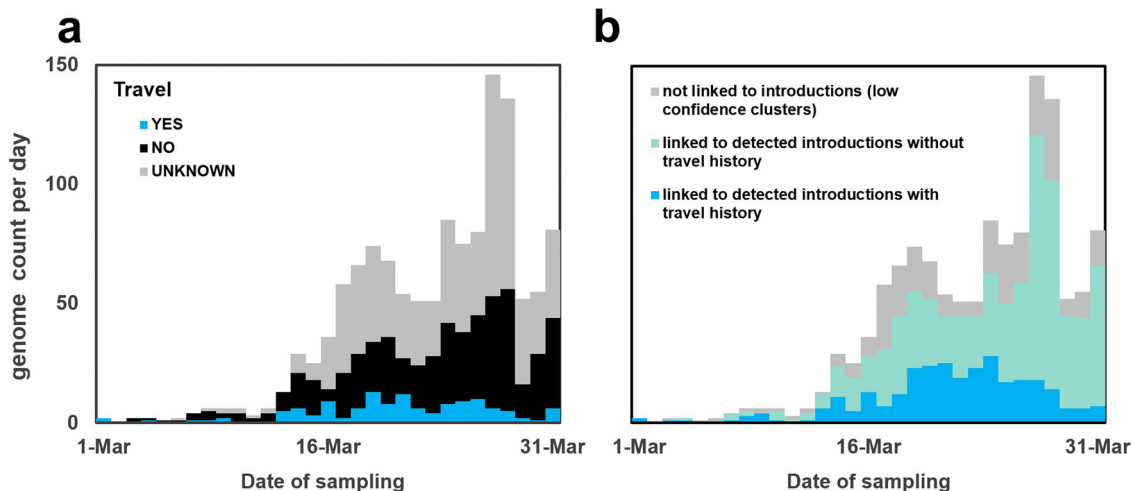
These analyses revealed a median time lag of 14 days (range = 3 and 53 days; Fig. 6) between the time to the PMRCA (TPMRCA; see Methods section for details), representing the earliest an introduction could have occurred, and surveilled genomes. In particular, the TPMRCA of most clades (~58%)



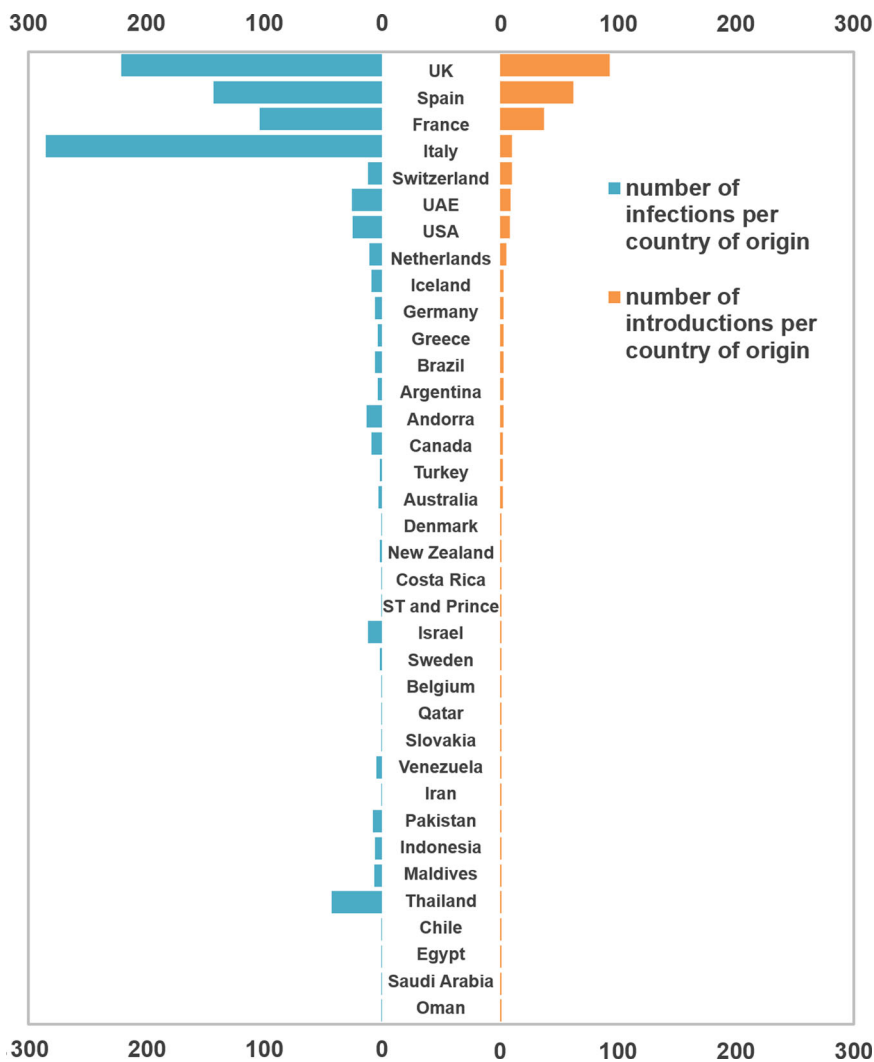
**Fig. 1 Overview of the COVID-19 confirmed cases and SARS-CoV-2 genome sequencing sampling in Portugal during the early phase of the pandemic.** **a** Daily reported COVID-19 confirmed cases in Portugal and key milestones during the early phase of the pandemic (source: Directorate-General of Health, <https://covid19.min-saude.pt/relatorio-de-situacao/>). **b** Area plots (right y-axis) reflect the cumulative total number of COVID-19 confirmed cases (gray) and SARS-CoV-2 genome sequences (blue) reported/obtained in Portugal, until 31 March 2020. The blue line (left y-axis) displays the "sequencing sampling", i.e. the proportion of confirmed cases with SARS-CoV-2 genome data during the same period. **c** SARS-CoV-2 sequencing sampling generated by the Health Administration region until 31 March 2020 (circles are proportional to the number of confirmed cases by Region, with the blue representing the proportion of samples with SARS-CoV-2 genome data). NOR Northern region, CEN Central region, LX+TV Lisbon and Tagus Valley region, ALN Alentejo, ALG Algarve, AZO Autonomous Region of Azores, MAD Autonomous Region of Madeira. **d** Area plots reflect the cumulative total number of COVID-19 confirmed cases (gray) and SARS-CoV-2 genome sequences (blue) reported by country between 22 January and 31 March 2020 for the top 10 countries with the highest number of genomes with collection date until 31 March 2020 (available on GISAID by 8th August 2020). A log<sub>10</sub> scale y-axis was used for visualization purposes. Countries are ordered according to the date of the first reported COVID-19 case.

occurred between the last week of February and the first week of March (Supplementary Data 8). The temporal reconstruction estimated the earliest introduction of lineage A into Portugal to have occurred as early as 7 March 2020 [95% Highest Posterior Density (HPD): 4 March 2020–10 March 2020], from Spain to the Lisbon & Tagus Valley region. The earliest seeding event for lineage B was estimated to have taken place on 2 February 2020 (95% HPD: 30 January 2020–5 February 2020) via the United Kingdom into the North region. Of note, 108 out of the 277 introductions were seeded by Nextstrain 20B clade viruses.

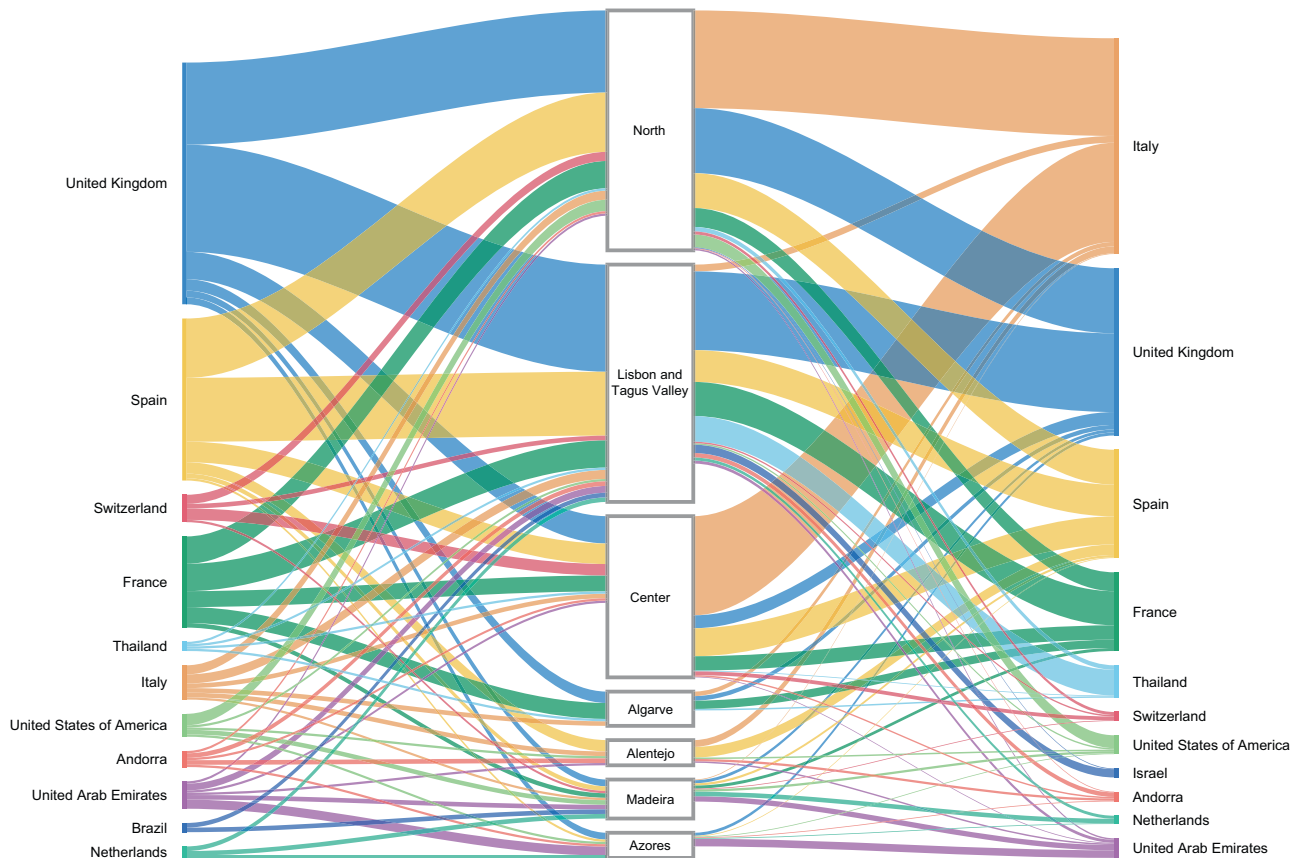
Because of both the high abundance of 20B sequences in the dataset and the low genetic diversity within this clade at the time (leading to polytomy-rich topologies), the TPMRCAs of these introductions might be less accurate in representing the importation date into Portugal. Notwithstanding, our data generally indicates that most introductions were seeded before main public health interventions and when the cumulative number of detected cases was still very low (Figs. 1 and 6), suggesting a period of several weeks where undetected transmission likely occurred (Fig. 6).



**Fig. 2 Overview of sequences linked to detected introductions, stratified by travel history data.** **a** Histogram of COVID-19 confirmed cases with SARS-CoV-2 genome sequencing data stratified by travel history. **b** Histogram with sequence counts stratified by respective reconstruction linked with clades that include sequences from cases with (blue) or without (green) known travel history. Sequences belonging to BEAST clades with low topology posterior probability are shown in gray.



**Fig. 3 Number and size of SARS-CoV-2 introductions per country.** Bar plots represent the number of introductions by country of origin (orange) and respective total number of cases (blue), inferences are derived from the phylodynamic reconstruction.



**Fig. 4 Sankey plots summarizing the viral flow into Portuguese Health Administration regions.** Left. The plot shows the relative number of transitions between countries of origin and different regions in Portugal. For summaries that show all transitions to and from all connected locations, we refer to the Supplementary Data 4. Right. The plot shows the relative number of infections generated by introductions from countries of origin and different regions in Portugal. For summaries that show the size of all introductions to and from all connected locations, we refer to the Supplementary Data 5. For both panels, for graphics simplicity, we present the eleven countries linked with most introductions (left) and relative number of infections generated (right), estimated across phylogenies inferred for the whole dataset, i.e., including both Pango lineages A and B (thin - low proportion of number/size of viral introductions attributed to this source; thick - high proportion of number/size of viral introductions attributed to this source). We note that there is no temporal order for the transitions involved.

**Discussion**

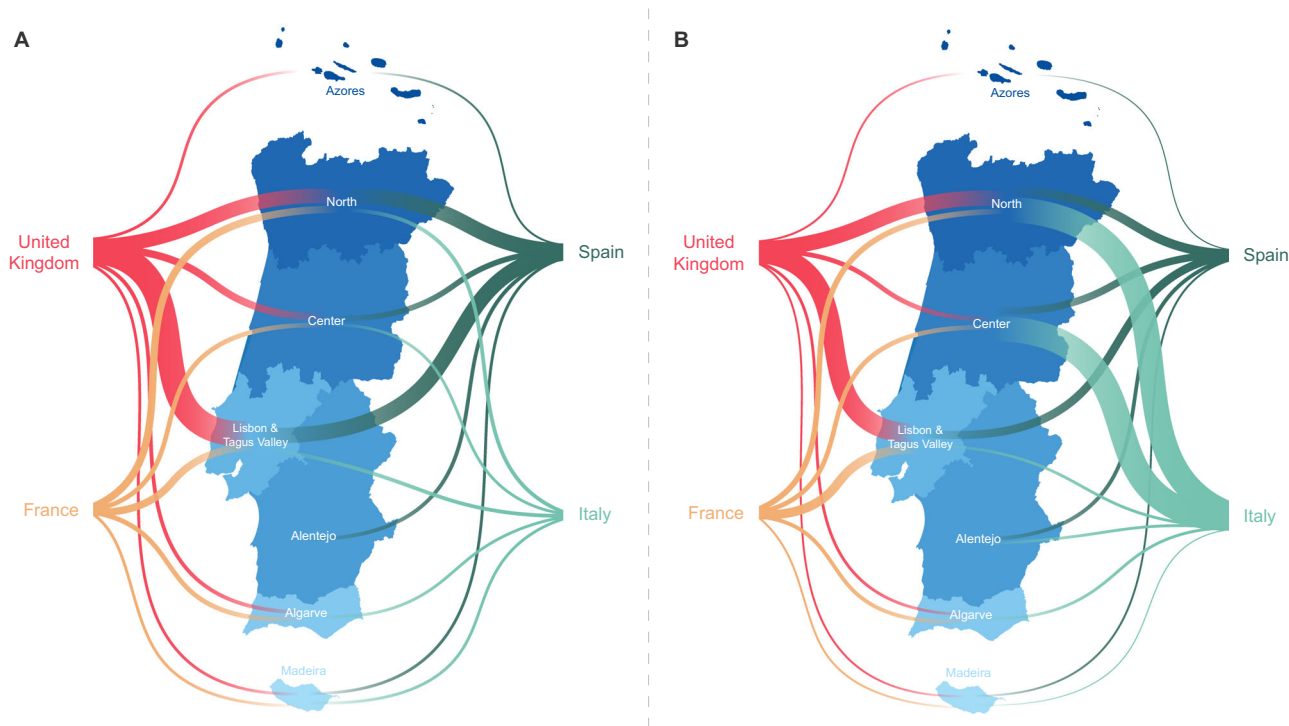
The present study provides a comprehensive description of the early establishment of the COVID-19 pandemic in Portugal based on a genomic epidemiology and phylodynamics approach. We detected multiple independent SARS-CoV-2 introductions, mostly from European countries (namely the UK, Spain, France, Italy, and Switzerland), which was broadly consistent with the available travel history data, as well as with the countries with most frequent flight connectivity with Portugal (INE, 2020) and/or with the highest number of Portuguese immigrants<sup>30,31</sup>.

The genomic surveillance efforts in Portugal culminated with the generation of 1275 SARS-CoV-2 whole genomes, which represents 15.5% of all confirmed cases in March, making Portugal the country generating the 5th highest number of SARS-CoV-2 genomic sequences during the early COVID-19 pandemic. In particular, our data uncovered the importation of extensive SARS-CoV-2 genomic diversity during the early epidemic in Portugal, as supported by the identification of at least 277 seeding events, covering all Nextstrain/GISAID clades and dozens of Pango lineages. Despite the vast genetic diversity introduced, 32% of the 1275 analyzed sequences could be linked to just ten out of the 277 detected introductions. In particular, a single introduction (from Italy) of a Spike Y839 variant represented about 20% of all sampled genomes collected until the end of March<sup>15</sup>. Moreover, 56% (155/277) of BEAST clades involve one single sequence

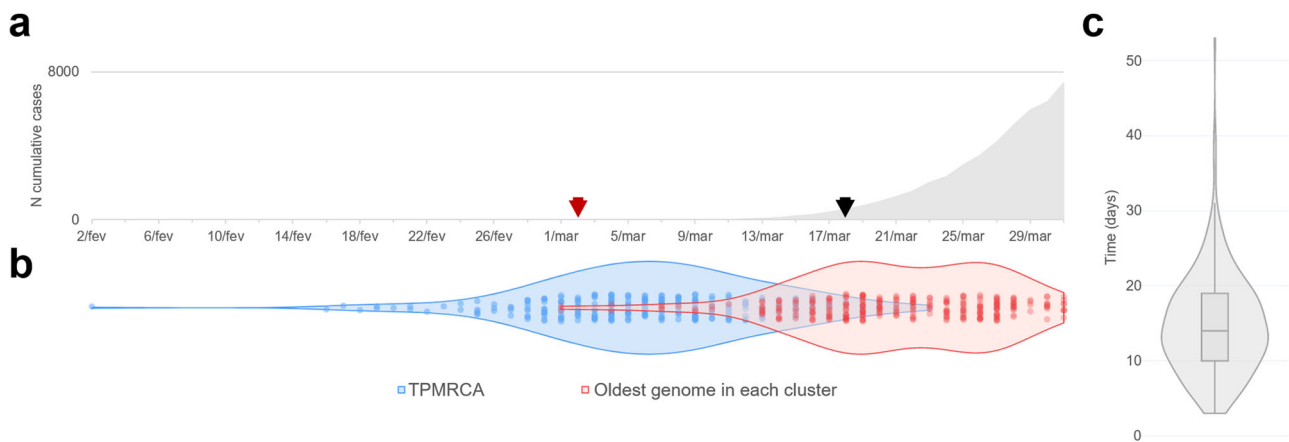
(singletons), thus suggesting that most introductions have not seeded substantial local transmission.

In general, these data suggest that the implemented preventive and early control measures have been successful in mitigating the establishment of community transmission from most SARS-CoV-2 independent introductions. On the other hand, it also highlights that the few introductions not captured by public health surveillance and control can still seed large community transmission events and give rise to a significant number of cases. This underlies the challenges in defining a public health strategy aimed at preventing sustained community transmission while maintaining open borders and global connectivity. Portugal presents a geopolitical and demographic context involving high circulation of both migrant workers and tourists, and close international networks with multiple countries from different latitudes (particularly from Europe, South America, and Africa). More stringent border restrictions, including flight and land borders closure, were implemented after March 10, when 74% (205/277) of the detected introductions were estimated to have already occurred. Despite the delay in the epidemic start in Portugal in comparison with other European countries, which certainly contributed to a more favorable situation during the pandemic’s first epidemic wave, the earlier implementation of measures (close borders, travelers testing, etc.) could have largely minimized the number of introductions and subsequent virus expansion.





**Fig. 5 Map summarizing the viral flow into Portuguese Health Administration regions.** **A** The plot shows the relative number of transitions between the four countries of origin linked with most introductions to different regions in Portugal. For summaries that show all transitions to and from all connected locations, we refer to the Supplementary Data 4. **B** The plot shows the relative number of infections generated by introductions from the four countries of origin linked with most relative number of infections generated in different regions in Portugal. For summaries that show the size of all introductions to and from all connected locations, we refer to the Supplementary Data 5. For a broader overview of the 11 countries linked with most introductions and relative number of infections generated, we also refer to Fig. 4. For both panels, introductions (left) and relative number of infections generated (right) were estimated across phylogenies inferred for the whole dataset, i.e., including both Pango lineages A and B (thin - low proportion of number/size of viral introductions attributed to this source; thick - high proportion of number/size of viral introductions attributed to this source). We note that there is no temporal order for the transitions involved.



**Fig. 6 Cryptic transmission of SARS-CoV-2 in Portugal revealed by genomic epidemiology.** **a** Number of cumulative cases over time (red and black arrows indicate the date of notification of the first COVID-19 case detected in Portugal and the start of the general lockdown, respectively) **b** Violin plots represent the date of sample collection of the oldest genome in a Portuguese clade (red) and the times for the parent nodes of the most recent common ancestors (TPMRCA in blue) for each of the 277 inferred introductions. **c** Violin and box plots depict the time lag between the introduction and the first surveilled genome for all 277 clusters.

The detection lag observed between the estimated time of the most recent common ancestor and the collection date of the earliest sample of each BEAST clade denotes that cryptic transmission might have occurred to some extent. Although most introductions were estimated to have occurred during the last week of February and, especially, during the first week of March

2020, SARS-CoV-2 was silently circulating in Portugal a few weeks before the first confirmed local cases on 2 March 2020. For instance, 12 out of 19 sequences collected during the first week of the epidemic were obtained from patients with no travel history, including patients linked to the introduction leading to the highest number of cases during the first wave. These results are



reasonable epidemiologically given the period needed for the diagnostics infrastructure to be set up. Also, the period of cryptic transmission reported in this study for Portugal is within the estimated for other countries during the early pandemic period<sup>32,33</sup>.

Phylogenetic modeling relies on the accumulation of genetic diversity over time for the estimation of the evolutionary rate and timing of other relevant events. Being a recently emerged pathogen with a relatively low substitution rate (in comparison with other RNA viruses) and a large genome, it is challenging to study the early dynamics of SARS-CoV-2, as a large proportion of sequences collected across time and space are very closely related genetically, thus complicating the reconstruction of the phylogenetic topology. To address this, we opted to focus on a subset of the results with higher topological support. Furthermore, due to the computational burden required for phylodynamic analysis, it is unfeasible to include all available genomic sequences. One limitation of this and other studies trying to infer the origin of introductions is the sequencing sampling bias by country. For this reason, we resorted to filtering and subsampling strategies that aim at reducing the number of genomes that do not significantly contribute to the evolutionary or phylogeographic reconstructions. Despite being less pronounced, there were still discrepancies in the number of genomes included across countries. For instance, it is very likely that the number of introductions via the UK (the country generating the highest volume of sequences) is overestimated, while the number of introductions from countries with none or few available genome sequences is underestimated. However, the integration of travel history during the phylodynamic inferences allows for a more accurate reconstruction of the spatial pathways from both over and undersampled locations. Additionally, using this recently developed model<sup>13</sup> provides insight into the genetic diversity circulating in countries for which genomic surveillance is still lacking. This makes the present work among the first to apply this novel model, which allowed us to gain insight into the early circulating diversity in countries like Saint Thomas and Prince, for which (as of 15 February 2021) there are no sequences available, and other undersampled countries such as the United Arab Emirates and Qatar.

Overall, our findings highlight the use of genomic data to trace the introduction and spread of an emerging virus, showing the need of systematic, continuous, and geographically representative genomic surveillance to detect and monitor the emergence and dissemination of biologically and/or epidemiologically relevant variants. Together with open data sharing, the timely generation of SARS-CoV-2 genomic data has been shown to be an invaluable tool to guide national and international public health authorities towards the identification and control of highly transmissible and/or immune evading variants<sup>34</sup>.

In this context, by laying the foundation of the genomic epidemiology of SARS-CoV-2 in Portugal involving an unprecedented collaborative effort at national and international levels, this work and concomitant capacity building, was pivotal for Portugal's response to the current and upcoming needs of genomic-informed surveillance and epidemiology of COVID-19, as strongly recommended by ECDC and WHO<sup>10,35,36</sup>.

### Data availability

SARS-CoV-2 genome sequences generated in this study were uploaded to the GISAID database (<https://www.gisaid.org/>). Accession numbers can be found in Supplementary material (Supplementary Data 1). Source data for the main figures in the manuscript can be accessed as: (i) Fig. 1. Daily reported COVID-19 confirmed cases in Portugal (<https://covid19.min-saude.pt/relatorio-de-situacao/>) (Fig. 1a–c); SARS-CoV-2 genome sequences reported/obtained in Portugal, until 31 March 2020 (Supplementary Data 1; <https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>) (Fig. 1b, c); Cumulative total number of COVID-19 confirmed cases and SARS-CoV-2 genome sequences reported by country between January 22 and March 31 (Supplementary Data 2 and 3; [https://github.com/CSSEGISandData/COVID-19/blob/master/csse\\_covid\\_19\\_data/](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/)

<https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>); (ii) Fig. 2. Sequences linked to detected introductions, stratified by travel history data (<https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>); (iii) Fig. 3. Number and size of SARS-CoV-2 introductions per country (<https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e> and Supplementary Data 8); Figs. 4e and 5. Relative number of transitions between countries of origin and different regions in Portugal (Supplementary Data 4; (<https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>)) and relative number of infections generated by introductions from countries of origin and different regions in Portugal. (Supplementary Data 5; (<https://microreact.org/project/cM6KURnU7rUpqdAnBq5DAf/a2d3840e>)); Fig. 6. Daily reported COVID-19 confirmed cases in Portugal (<https://covid19.min-saude.pt/relatorio-de-situacao/>) and size, mean date estimates (and respective 95% highest posterior density intervals) of parent nodes of the MRCA and the earliest date of sample collection for each BEAST cluster (Supplementary Data 8).

Received: 21 May 2021; Accepted: 5 January 2022;

Published online: 28 January 2022

### References

- Wu, F. et al. Author Correction: a new coronavirus associated with human respiratory disease in China. *Nature* **580**, E7 (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- World Health Organization (WHO). *Weekly operational update on COVID-19* [https://www.who.int/docs/default-source/coronaviruse/weekly-updates/wou\\_2021\\_29-mar\\_project\\_cleared.pdf?sfvrsn=c0cd7458\\_3&download=true](https://www.who.int/docs/default-source/coronaviruse/weekly-updates/wou_2021_29-mar_project_cleared.pdf?sfvrsn=c0cd7458_3&download=true) (2021).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* **20**, 533–534 (2020). Erratum in: *Lancet Infect Dis.* 2020 Sep:e215. PMID: 32087114; PMCID: PMC7159018. <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>.
- Alm, E. et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill.* **25**, pii=2001410 (2020).
- Gámbaro, F. et al. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Euro Surveill.* **25**, 2001200 (2020).
- du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
- Fauver, J. R. et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell.* **181**, 990–996.e5 (2020). Epub 2020 May 7.
- Lemieux, J. E. et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science.* **371**, eabe3261 (2021). Epub 2020 Dec 10. PMID: 33303686; PMCID: PMC7857412.
- Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications/i/item/9789240018440>.
- Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
- Argimón, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* **2**, e000093 (2016).
- Lemey, P. et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat. Commun.* **11**, 5110 (2020). PMID: 33037213; PMCID: PMC7547076.
- Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
- Borges, V. et al. Massive dissemination of a SARS-CoV-2 Spike Y839 variant in Portugal. *Emerg. Microbes Infect.* **2**, 1–58 (2020).
- Borges, V., Pinheiro, M., Pechirra, P., Guiomar, R. & Gomes, J. P. INSAFLU: an automated open web-based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based surveillance. *Genome Med.* **10**, 46 (2018).
- Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
- Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
- Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
- Hong, S. L. et al. In search of covariates of HIV-1 subtype B spread in the United States—a cautionary tale of large-scale Bayesian phylogeography. *Viruses* **12**, 182 (2020).

22. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
23. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
24. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23**, 7–9 (2006).
25. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
26. Lauer, S. A. et al. The incubation period of Coronavirus Disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
27. Edwards, C. J. et al. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* **21**, 1251–1258 (2011).
28. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
29. Nadeau, S. A., Vaughan, T. G., Scire, J., Huisman, J. S. & Stadler, T. The origin and early spread of SARS-CoV-2 in Europe. *Proc. Natl Acad. Sci. USA* **118**, e2012008118 (2021).
30. United Nations, Department of Economic and Social Affairs. Population Division (2019). *International Migrant Stock 2019* (United Nations database, POP/DB/MIG/Stock/Rev.2019).
31. Alto Comissariado para as Migrações (ACM). (n.d.). *Saber mais sobre as migrações portuguesas*. Available at: <https://www.acm.gov.pt/-/saber-mais-sobre-as-migracoes-portuguesas-#>.
32. Bedford, T. et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*. **370**, 571–575 (2020).
33. Davis J. T. et al. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature* <https://doi.org/10.1038/s41586-021-04130-w> (2021).
34. European Centre for Disease Prevention and Control (ECDC). Risk related to spread of new SARS-CoV-2 variants of concern in the EU/EEA – 29 December 2020 (ECDC, 2020). <https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-risk-related-to-spread-of-new-SARS-CoV-2-variants-EU-EEA.pdf>.
35. World Health Organization. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. Geneva: Licence: CC BY-NC-SA 3.0 <https://apps.who.int/iris/handle/10665/338480> (2021).
36. European Centre for Disease Prevention and Control (ECDC). Sequencing of SARS-CoV-2: first update. <https://www.ecdc.europa.eu/sites/default/files/documents/Sequencing-of-SARS-CoV-2-first-update.pdf> (ECDC, 2021).

## Acknowledgements

We gratefully acknowledge to Sara Hill and Nuno Faria (University of Oxford) and Joshua Quick and Nick Loman (University of Birmingham) for kindly providing us with the initial sets of Artic Network primers for NGS; Rafael Mamede (MRamirez team, IMM, Lisbon) for developing and sharing a bioinformatics script for sequence curation (<https://github.com/rfm-targa/BioinfUtils>); Philippe Lemey (KU Leuven) for providing guidance on the implementation of the phylodynamic models; Joshua L. Cherry (National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health) for providing guidance with the subsampling strategies; and all authors, originating and submitting laboratories who have contributed genome data on GISAID (<https://www.gisaid.org/>) on which part of this research is based. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government. This study is co-funded by Fundação para a Ciência e Tec-

nologia and Agência de Investigação Clínica e Inovação Biomédica (234\_596874175) on behalf of the Research 4 COVID-19 call. Some infrastructural resources used in this study come from the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia (FCT).

## Author contributions

V.B., J.I., and J.P.G. designed the study. V.B. and J.I. performed the pre-sequencing wet-lab procedures, genome assembly, curation and classification, and genetic and epidemiological data curation and integration. H.C.M. performed metadata collection and curation. H.M. supported data divulgation by setting up the public website. S.D., L.V., R.L., and I.G. performed and/or supervised the wet-lab sequencing procedures. N.S.T. conceived and performed the Bayesian phylodynamic analyses. N.S.T. retrieved the external genetic data and assembled the datasets. The results were analyzed and interpreted by V.B., J.I., and N.S.T.; V.B., J.I., N.S.T., and J.P.G. wrote and reviewed the manuscript. All authors read and approved the final manuscript. All authors from the Portuguese network for SARS-CoV-2 genomics (Consortium) contributed to laboratory diagnosis of SARS-CoV-2, sample and data collection, and/or provided advice on epidemiology and bioinformatics data analysis. R.G. supported sample collection and pre-sequencing wet-lab procedures. J.P.G. coordinated the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-022-00072-0>.

**Correspondence** and requests for materials should be addressed to João Paulo Gomes.

**Peer review information** *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

## Portuguese network for SARS-CoV-2 genomics (Consortium)

Agostinho José S. Lira<sup>8</sup>, Aida M. Sousa Fernandes<sup>9</sup>, Alexandra Estrada<sup>10</sup>, Alexandra Nunes<sup>11</sup>, Alfredo Rodrigues<sup>12</sup>, Ana Caldas<sup>13</sup>, Ana Constança<sup>14</sup>, Ana Margarida Henriques<sup>15</sup>, Ana Miguel Matos<sup>16</sup>, Ana Oliveira<sup>17</sup>, Ana Paula Dias<sup>18</sup>, Ana Pelerito<sup>11</sup>, Ana Rita Couto<sup>19</sup>, Anabela Vilares<sup>11</sup>, António Albuquerque<sup>20</sup>, Baltazar Nunes<sup>11</sup>, Bruna R. Gouveia<sup>21</sup>, Carina de Fátima Rodrigues<sup>22</sup>, Carla Feliciano<sup>11</sup>, Carla Roque<sup>11</sup>,

Carlos Cardoso<sup>23</sup>, Carlos Sousa<sup>24</sup>, Cathy Paulino<sup>5</sup>, Célia Rodrigues Bettencourt<sup>11</sup>, Claudia C. Branco<sup>25</sup>, Cláudia Nunes dos Santos<sup>26</sup>, Conceição Godinho<sup>27</sup>, Constantino P. Caetano<sup>11</sup>, Cristina Correia<sup>11</sup>, Cristina Toscano<sup>18</sup>, Cristina Veríssimo<sup>11</sup>, Daniela Silva<sup>28</sup>, Diana Patrícia Pinto da Silva<sup>29</sup>, Eliana Costa<sup>30</sup>, Elizabeth Pádua<sup>11</sup>, Fátima Martins<sup>11</sup>, Fátima Vale<sup>31</sup>, Fernanda Vilarinho<sup>11</sup>, Fernando Branca<sup>10</sup>, Filomena Caldeira<sup>32</sup>, Filomena Lacerda<sup>8</sup>, Francisca Rocha<sup>33</sup>, Graça Andrade<sup>34</sup>, Helena Ribeiro<sup>35</sup>, Helena Rodrigues<sup>33</sup>, Herberto Jesus<sup>36</sup>, Hugo Sousa<sup>37</sup>, Idalina Ferreira<sup>11</sup>, Inês Baldaque<sup>37</sup>, Inês Costa<sup>11</sup>, Inês Gomes<sup>28</sup>, Inna Slobidnyk<sup>38</sup>, Isabel Albergaria<sup>13</sup>, Isabel Dias<sup>27</sup>, Isabel Fernandes<sup>27</sup>, Isabel Lopes de Carvalho<sup>11</sup>, Ivone Água-Doce<sup>11</sup>, Jácome Bruges Armas<sup>19</sup>, Joana Ramos<sup>31</sup>, João Carlos Sousa<sup>39</sup>, João Costa<sup>5</sup>, João Dias<sup>18</sup>, João Rodrigues<sup>21</sup>, João Sobral<sup>5</sup>, Jorge Machado<sup>11</sup>, Jorge Meneses<sup>8</sup>, José Alves<sup>34</sup>, José Vicente Constantino<sup>11</sup>, Laura Brum<sup>40</sup>, Leonor Silveira<sup>11</sup>, Líbia Zé-Zé<sup>11</sup>, Lidia Santos<sup>27</sup>, Ludivina Freitas<sup>34</sup>, Luís Silva<sup>35</sup>, Luisa Mota-Vieira<sup>25</sup>, Lurdes Lopes<sup>27</sup>, Lurdes Monteiro<sup>41</sup>, Márcia Faria<sup>11</sup>, Margarida Farinha<sup>13</sup>, Margarida Vaz<sup>11</sup>, Maria Alice Pinto<sup>22</sup>, Maria Ana Pessanha<sup>18</sup>, Maria Beatriz Tomaz<sup>12</sup>, Maria Calle Vellés<sup>42</sup>, Maria da Graça Maciel de Soveral<sup>43</sup>, Maria Helena Ramos<sup>14</sup>, Maria Isabel Veiga<sup>39</sup>, Maria João Gargate<sup>11</sup>, Maria João Peres<sup>44</sup>, Maria José Borrego<sup>11</sup>, Maria Matos Figueiredo<sup>8</sup>, Mariana Martins<sup>45</sup>, Mariana Viana<sup>42</sup>, Maurício Melim<sup>36</sup>, Miguel Babarro Jorrete<sup>43</sup>, Miguel Fevereiro<sup>15</sup>, Miguel Pinheiro<sup>46</sup>, Mónica Oleastro<sup>11</sup>, Nair Seixas<sup>8</sup>, Nelson Ventura<sup>31</sup>, Nuno Verdasca<sup>11</sup>, Olga Costa<sup>27</sup>, Patrícia Barros<sup>11</sup>, Patricia Fonseca<sup>31</sup>, Patricia Miguel<sup>27</sup>, Paula Bajanca-Lavado<sup>11</sup>, Paula Branquinho<sup>27</sup>, Paula Palminha<sup>11</sup>, Paula Soares<sup>27</sup>, Paula Valente<sup>47</sup>, Paulo Leandro<sup>8</sup>, Paulo Pereira<sup>26</sup>, Pedro Cardoso<sup>23</sup>, Pedro Pechirra<sup>11</sup>, Pedro Ramos<sup>48</sup>, Raquel Neves<sup>11</sup>, Raquel Rocha<sup>11</sup>, Raquel Rodrigues<sup>11</sup>, Raquel Sabino<sup>11</sup>, Regina Sá<sup>17</sup>, Ricardo Filipe Romão Ferreira<sup>29</sup>, Ricardo Rodrigues<sup>45</sup>, Rita C. Veloso<sup>25</sup>, Rita Cordeiro<sup>11</sup>, Rita Côrte-Real<sup>27</sup>, Rita de Sousa<sup>11</sup>, Rita Gralha<sup>31</sup>, Rita Macedo<sup>11</sup>, Rita Matos<sup>11</sup>, Rita Rodrigues<sup>35</sup>, Sandra Paulo<sup>45</sup>, Sara Sousa<sup>30</sup>, Sílvia Lopo<sup>11</sup>, Sónia Marta Santos Magalhães<sup>49</sup>, Sónia Rodrigues<sup>27</sup>, Sónia Silva<sup>11</sup>, Susana Ladeiro<sup>5</sup>, Susana Martins<sup>11</sup>, Susana Silva<sup>11</sup>, Teresa Salvado<sup>35</sup>, Tiago Luís<sup>15</sup>, Valquíria Alves<sup>20</sup> & Vera Manageiro<sup>11</sup>

<sup>8</sup>Centro Hospitalar de Vila Nova de Gaia e Espinho, Vila Nova de Gaia, Portugal. <sup>9</sup>Laboratório Regional de Saúde Pública Dr<sup>a</sup> Laura Ayres - ARS Algarve, Almancil, Portugal. <sup>10</sup>Serviço de Patologia Clínica, Hospital de Braga, Braga, Portugal. <sup>11</sup>Instituto Nacional de Saúde Dr Ricardo Jorge (INSA), Lisboa, Portugal. <sup>12</sup>Beatriz Godinho Saúde, Leiria, Portugal. <sup>13</sup>Serviço de Patologia Clínica, Centro Hospitalar Tondela-Viseu, Viseu, Portugal. <sup>14</sup>Serviço de Microbiologia, Centro Hospitalar do Porto, Porto, Portugal. <sup>15</sup>Instituto Nacional de Investigação Agrária e Veterinária, Lisboa, Portugal. <sup>16</sup>Laboratório de Análises Clínicas da Universidade de Coimbra, Coimbra, Portugal. <sup>17</sup>ACES de Baixo Vouga, Administração Regional de Saúde do Centro, Aveiro, Portugal. <sup>18</sup>Laboratório de Microbiologia e Biologia Molecular do Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal. <sup>19</sup>Serviço Especializado de Epidemiologia e Biologia Molecular, Hospital de Santo Espírito, Ilha Terceira, Açores Angra do Heroísmo, Portugal. <sup>20</sup>Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal. <sup>21</sup>Interactive Technologies Institute - LARSyS, Funchal, Portugal. <sup>22</sup>Centro de Investigação de Montanha, Instituto Politécnico de Bragança, Bragança, Portugal. <sup>23</sup>Laboratório de Análises Clínicas Dr Joaquim Chaves, Lisboa, Portugal. <sup>24</sup>Laboratório De Biologia Molecular da Unilabs, Porto, Portugal. <sup>25</sup>Unidade de Genética e Patologia Moleculares, Hospital do Divino Espírito Santo de Ponta Delgada, Ilha de S. Miguel, Açores, Ponta Delgada, Portugal. <sup>26</sup>Centro de Estudos de Doenças Crónicas, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal. <sup>27</sup>Laboratório Biologia Molecular- Serviço de Patologia Clínica, Centro Hospitalar Universitário Lisboa Central, Lisboa, Portugal. <sup>28</sup>ALS Controlvet, Tondela, Portugal. <sup>29</sup>Centro Médico da Praça, São João da Madeira, Portugal. <sup>30</sup>Serviço de Patologia Clínica, Centro Hospitalar de Trás-os-Montes e Alto Douro, Vila Real, Portugal. <sup>31</sup>Serviço de Patologia Clínica, Unidade Local de Saúde da Guarda, Guarda, Portugal. <sup>32</sup>Hospital Espírito Santo, Évora, Portugal. <sup>33</sup>Laboratório Dra Helena Rodrigues, Valença, Portugal. <sup>34</sup>Serviço de Patologia Clínica - Hospital Dr. Nélio Mendonça - SESARAM, Funchal, Portugal. <sup>35</sup>Serviço de Patologia Clínica, Hospital de Vila Franca de Xira, Vila Franca de Xira, Portugal. <sup>36</sup>Instituto de Administração da Saúde da Madeira, Funchal, Portugal. <sup>37</sup>Serviço de Virologia, Instituto Português de Oncologia do Porto, Porto, Portugal. <sup>38</sup>Serviço de Patologia Clínica - Unidade Local de Saúde Litoral Alentejano, Santiago do Cacém, Portugal. <sup>39</sup>Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga, Portugal. <sup>40</sup>Synlab, Lisboa, Portugal. <sup>41</sup>Secção de Patologia Molecular, Synlab, Lisboa, Portugal. <sup>42</sup>Centro Hospitalar Tâmega e Sousa, Penafiel, Portugal. <sup>43</sup>Serviço de Imunohemoterapia, Unidade Local de Saúde do Alto Minho, Viana do Castelo, Portugal. <sup>44</sup>Laboratório de Imunologia e Biologia Molecular, Centro Hospitalar de Setúbal, Setúbal, Portugal. <sup>45</sup>Serviço de Patologia Clínica da Unidade Local de Saúde de Castelo Branco, Castelo Branco, Portugal. <sup>46</sup>iBiMED/Universidade de Aveiro, Aveiro, Portugal. <sup>47</sup>Departamento de Saúde Pública e Planeamento, ARS Alentejo, Évora, Portugal. <sup>48</sup>Secretaria Regional de Saúde e Proteção Civil - Governo Regional da Madeira, Funchal, Portugal. <sup>49</sup>Hospital Agostinho Ribeiro-Felgueiras, Felgueiras, Portugal.