The first wave of the Spanish COVID-19 epidemic was associated with early introductions and fast spread of a dominating genetic variant

Mariana G. López[1,#], Álvaro Chiner-Oms[1,#], Darío García de Viedma[2,3,4], Paula Ruiz-Rodriguez[5], Maria Alma Bracho[6,7], Irving Cancino-Muñoz[1], Giuseppe D'Auria[7,8], Griselda de Marco[8], Neris García-González[6], Galo Adrian Goig[9], Inmaculada Gómez-Navarro[1], Santiago Jiménez-Serrano[1], Llúcia Martinez-Priego[8], Paula Ruiz-Hueso[8], Lidia Ruiz-Roldán[6], Manuela Torres-Puente[1], Juan Alberola[10,11,12], Eliseo Albert[13], Maitane Aranzamendi Zaldumbide[14,15], María Pilar Bea-Escudero[16], Jose Antonio Boga[17,18], Antoni E. Bordoy[19], Andrés Canut-Blasco[20], Ana Carvajal[21], Gustavo Cilla Eguiluz[22], Maria Luz Cordón Rodríguez[20], José J. Costa-Alcalde[23], María de Toro[16], Inmaculada de Toro Peinado[24], Jose Luis del Pozo[25], Sebastián Duchêne[26], Jovita Fernández-Pinero[27], Begoña Fuster Escrivá[28,29], Concepción Gimeno Cardona[28], Verónica González Galán[30,31], Nieves Gonzalo Jiménez[32], Silvia Hernáez Crespo[20], Marta Herranz[2,3,4], José Antonio Lepe[30,31], José Luis López-Hontangas[33], Maria Ángeles Marcos[34], Vicente Martín[7,35], Elisa Martró[7,19], Ana Milagro Beamonte[36], Milagrosa Montes Ros[22], Rosario Moreno-Muñoz[37], David Navarro[13,29], José María Navarro-Marí[38,39], Anna Not[19], Antonio Oliver[40,41], Begoña Palop-Borrás[24], Mónica Parra Grande[42], Irene Pedrosa-Corral[38,39], Maria Carmen Perez Gonzalez[43], Laura Pérez-Lago[2,3], Luis Piñeiro Vázquez[22], Nuria Rabella[44,45,46], Jordi Reina[40], Antonio Rezusta[36,47,48], Lorena Robles Fonseca[42], Ángel Rodríguez-Villodres[30,31], Sara Sanbonmatsu-Gámez[38,39], Jon Sicilia[2,3], María Dolores Tirado Balaguer[37], Ignacio Torres[13], Alexander Tristancho[36,47], José María Marimón[22], Mireia Coscolla[5,*], Fernando González-Candelas[6,7,*], Iñaki Comas[1,7,*] on behalf of the SeqCOVID-SPAIN consortium.

[1]Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain
[2]Servicio de Microbiología Clínica y Enfermedades Infecciosas. Hospital General Universitario Gregorio Marañón, Madrid, Spain
[3]Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[4]CIBER Enfermedades Respiratorias (CIBERES)
[5]Instituto de Biología Integrativa de Sistemas, I2SysBio (CSIC-Universitat de València), Valencia, Spain
[6]Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia I2SysBio, Valencia, Spain
[7]CIBER in Epidemiology and Public Health (CIBERESP)
[8]Sequencing and Bioinformatics Service of the Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO-Public Health), Valencia, Spain
[9]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland

[10]Servicio de Microbiología, Hospital Universitario Doctor Peset, Valencia, Spain
[11]Conselleria de Sanitat i Consum, Generalitat Valenciana, Valencia, Spain
[12]Departamento de Microbiología, Facultad de Medicina, Universidad de Valencia, Valencia, Spain
[13]Microbiology Service, Hospital Clínico Universitario, INCLIVA Research Institute, Valencia, Spain
[14]Servicio de Microbiología, Hospital Universitario Cruces, Bilbao, Spain
[15]Grupo de Microbiología y Control de Infección Instituto de Investigación Sanitaria Biocruces Bizkaia, Spain
[16]Plataforma de Genómica y Bioinformática, Centro de Investigación Biomédica de La Rioja (CIBIR), Logroño, Spain
[17]Servicio de Microbiología, Hospital Universitario Central de Asturias, Oviedo, Spain
[18]Grupo de Microbiología Traslacional Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Spain
[19]Servicio de Microbiología, Laboratori Clínic Metropolitana Nord, Hospital Universitari Germans Trias i Pujol, Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Badalona, Barcelona, Spain
[20]Servicio de Microbiología, Hospital Universitario de Álava, Osakidetza-Servicio Vasco de Salud, Vitoria-Gasteiz, Álava, Spain
[21]Animal Health Department. Universidad de León, León, Spain
[22]Biodonostia, Osakidetza, Hospital Universitario Donostia, Servicio de Microbiología, San Sebastián, Spain
[23]Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain
[24]Servicio de Microbiología, Hospital Regional Universitario de Málaga, Málaga, Spain
[25]Servicio de Enfermedades Infecciosas y Microbiología Clínica. Clínica Universidad de Navarra, Pamplona, Spain
[26]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia
[27]Centro de Investigación en Sanidad Animal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, O.A., M.P. - INIA, Valdeolmos, Spain
[28]Servicio de Microbiología, Consorcio Hospital General Universitario de Valencia, Valencia, Spain
[29]Department of Microbiology, School of Medicine, University of Valencia, Valencia, Spain.
[30]Servicio de Microbiología UCEIMP, Sevilla, Spain
[31]Hospital Universitario Virgen del Rocío, Sevilla, Spain
[32]Servicio Microbiología, Departamento de Salud de Elche-Hospital General, Elche, Alicante, Spain
[33]Servicio de Microbiología, Hospital Universitario y Politécnico La Fe, Valencia, Spain
[34]Micobiology Department, Hospital Clinic i Provincial de Barcelona, Institute of Global Health of Barcelona (ISGlobal), Barcelona, Spain
[35]Research Group on Gene-Environment Interactions and Health. Institute of Biomedicine (IBIOMED). Universidad de León, León, Spain
[36]Servicio de Microbiología Clínica Hospital Universitario Miguel Servet, Zaragoza, Spain
[37]Hospital General Universitario de Castellón, Castellón, Spain

[38]Servicio de Microbiología. Hospital Universitario Virgen de las Nieves, Granada, Spain
[39]Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria ibs, Granada, Spain
[40]Servicio de Microbiología, Hospital Universitario Son Espases, Palma de Mallorca, Spain
[41]Instituto de Investigación Sanitaria de las Islas Baleares, Spain
[42]Hospital General Universitario de Albacete, Spain
[43]Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain
[44]Servei de Microbiologia, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[45]CREPIMC. Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain
[46]Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Cerdanyola
[47]Instituto de Investigación Sanitaria de Aragón, Centro de Investigación Biomédica de Aragón (CIBA), Zaragoza, Spain
[48]Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain

[#] Contributed equally
* Corresponding authors

ABSTRACT

The COVID-19 pandemic has shaken the world since the beginning of 2020. Spain is among the European countries with the highest incidence of the disease during the first pandemic wave. We established a multidisciplinar consortium to monitor and study the evolution of the epidemic, with the aim of contributing to decision making and stopping rapid spreading across the country. We present the results for 2170 sequences from the first wave of the SARS-Cov-2 epidemic in Spain and representing 12% of diagnosed cases until 14th March. This effort allows us to document at least 500 initial introductions, between early February-March from multiple international sources. Importantly, we document the early raise of two dominant genetic variants in Spain (Spanish Epidemic Clades), named SEC7 and SEC8, likely amplified by superspreading events. In sharp contrast to other non-Asian countries those two variants were closely related to the initial variants of SARS-CoV-2 described in Asia and represented 40% of the genome sequences analyzed. The two dominant SECs were widely spread across the country compared to other genetic variants with SEC8 reaching a 60% prevalence just before the lockdown. Employing Bayesian phylodynamic analysis, we inferred a reduction in the effective reproductive number of these two SECs from around 2.5 to below 0.5 after the implementation of strict public-health interventions in mid March. The effects of lockdown on the genetic variants of the virus are reflected in the general replacement of preexisting SECs by a new variant at the beginning of the summer season. Our results reveal a significant difference in the genetic makeup of the epidemic in Spain and support the effectiveness of lockdown measures in controlling virus spread even for the most successful genetic variants. Finally, earlier control of SEC7 and particularly SEC8 might have reduced the incidence and impact of COVID-19 in our country.

MAIN

The new coronavirus disease (COVID-19) caused by SARS-CoV-2 emerged in China in October/November 2019[1] and by the end of March of 2020 it was present in most countries of the world. The World Health Organization declared the new disease as a pandemic on 11th March 2020. Spain suffered a severe epidemic with the first case notified on 29th January[2] and with an accumulated number of 261,584 cases by 1st July, including 29,965 fatalities. Furthermore, a nationwide seroprevalence study showed that only one in ten cases of infection by SARS-CoV-2 were diagnosed and declared in that period[3], suggesting that the total number of infections has been vastly underestimated. Spain ordered a series of non-pharmaceutical intervention measures including a general lockdown on 14th March[4], later applied by many other countries, and was successful in bending the curve by the end of May. Despite these measures, at least 30,000 individuals died during the first wave of the epidemic and a second wave of COVID-19 slowly started by the end of July 2020[5].

Despite the high incidence accumulated across the country some regions had significantly higher incidence than others. Genomic epidemiology and phylodynamics[6–8] offer a unique opportunity to understand the early events of the epidemic at the global, regional and local levels, to track the evolution of the epidemic after its initial stages and to quantify the impact of lockdown measures on the genetic variants of the virus. However, there are challenges and caveats that prevent the use of pathogen genomes as the sole source of interpretation. While there is now a large number of SARS-CoV-2 sequences deposited in the databases[9] there are still important unsampled areas of the world, including some that played an important role in the initial spread of the epidemic. In addition, the virus spreads faster than it evolves[10,11] which limits the resolution of phylogenetic and phylodynamic analysis[12]. Finally, despite important efforts by sequencing consortiums, only a fraction of the total number of infections has been sequenced. Nevertheless, genomic epidemiology has played an important role in understanding the global and local epidemiology of COVID-19[13–15].

After the pandemic was declared in Spain, we assembled the National Consortium of genomic epidemiology of SARS-CoV-2 (http://seqcovid.csic.es/). This established a unique network incorporating more than 50 hospitals and scientific institutions across the country to collect clinical samples and epidemiological information from COVID-19 cases. Here we present the results of this nation-wide effort. We were able to sequence 12% of the reported cases before the national lockdown, and 1% of the reported cases of the first wave (until 14th May), including samples of SARS-CoV-2 across Spain in the early months of the pandemic (February-May). Using a combination of pathogen genomics, phylogenetic tools, clinical and epidemiological data we have been able to dissect the very early events in the dispersion of SARS-CoV-2 throughout Spain as well the evolution of the virus during the exponential phase and after the lockdown. We document simultaneous introductions in the country from multiple sources. We show that up to 40% of cases were caused by two Spanish epidemic clades, named SEC7 and SEC8. Seven other Spanish epidemic clades were detected but their role was minor, probably because they were introduced relatively close to the lockdown and had no opportunities for a rapid exponential expansion as the

initial two clades had. In contrast to other European countries these SECs belong to early lineages in the epidemic (A in Pangolin, 19B in NextStrain). We also show that the reproductive number, R_e, of the most successful Spanish epidemic clades quickly declined after the implementation of lockdown measures and they were completely absent from samples taken in July-September. Our results suggest that the most successful variants were those associated with earlier introductions but also that their success may depend on the synergy between superspreading events and high mobility. These results also show the effectiveness of lockdown measures in controlling the virus spread and eliminating established successful epidemic clusters from circulation.

SARS-CoV-2 was introduced multiple times from multiple sources

Our dataset consists of 2,170 sequences from Spain, collected under ethical approval, from 25th February to 22nd June, coinciding with the initial phases of the COVID-19 pandemic in the country (Figure 1a). The most populated Spanish regions were sampled, resulting in a dataset with sequences representing 16 of the 17 administrative regions in which the country is divided (Figure 1b). 1,962 out of the 2,170 (90.4%) samples analyzed here have been sequenced by the SeqCOVID consortium, while the remaining 208 have been generated by independent laboratories and downloaded from GISAID[9] (Table S1). Spain displayed a particular viral population structure with a higher proportion of lineage A sequences compared to other European countries[16](Figure 1c). Strains from patients in Spain were more closely related with cases sequenced in China and were the most abundant during the first weeks of the Spanish epidemic. They were replaced by lineage B strains (Figure S1), which differ by at least 6-7 substitutions from lineage A and dominated the beginning of the pandemic in most European countries. In addition, we observed an heterogeneous distribution of the SARS-CoV-2 genetic diversity within Spain, both at the regional and local levels. For example, our analysis shows how viral diversity declined with geographic distance from a large urban outdoor like Valencia (see Supplementary Notes).

Similarly to other countries[17,18], phylogenomic analyses suggest the existence of multiple independent entries of the virus into Spain. To identify possible introductions we inspected the placement of Spanish viral samples in a global phylogeny constructed with more than 30,000 sequences (Figure 1d). Given the low genetic diversity of the virus, particularly at the beginning of the epidemic, we found most instances in which a Spanish sample is genetically identical to other variants circulating in the rest of the world. According to their phylogenetic placement, three different possibilities were considered for the phylogenetic position of Spanish sequences. A sequence was included in a ‘candidate transmission cluster’ when it was found in a monophyletic clade with other Spanish sequences; it was included in a ‘zero distance’ group when it grouped with other genetically identical Spanish sequences but also with other foreign sequences; and it was denoted as ‘unique’ when no matching sequence in the Spanish dataset was identified (see detailed definitions of the groups in Mat and Met and in Figure S2). We detected 224 ‘candidate transmission clusters’ comprising 827 sequences (~40% of the Spanish samples); 30 ‘zero-distance clusters’, comprising 831 sequences, and 513 ‘unique’ sequences (Figures S3). Next, we determined how many unique cases and clusters were compatible with an introduction before

the general lockdown. We detected that 191 groups (165 'candidate transmission clusters' plus 26 'zero distance clusters') and 328 unique sequences met this criteria, representing at least 519 independent introductions (distribution of dates in Figure 1e). This is probably an underestimation of the total number of entries because the number of sequences analyzed is a small subset of the total notified cases (Figure 1a). Phylogenetic analysis suggests that the most likely introductions of cases with a clear phylogenetic link (see Methods) came from Italy, the Netherlands, England, and Austria (accounting for ~23%, ~20%, ~13% and 12% of the cases for which a likely country of origin can be inferred, respectively) (Figure 1f). The observation that more than half of the introduction events detected are unique sequences illustrates the heterogeneous outcome after an introduction, as some events resulted in large epidemiological clusters, and others disappeared leaving almost no trace. A clear example is the first described death in Spain for which we have generated a partial sequence and who was infected in Nepal but who did not generate any identifiable secondary cases in our dataset.

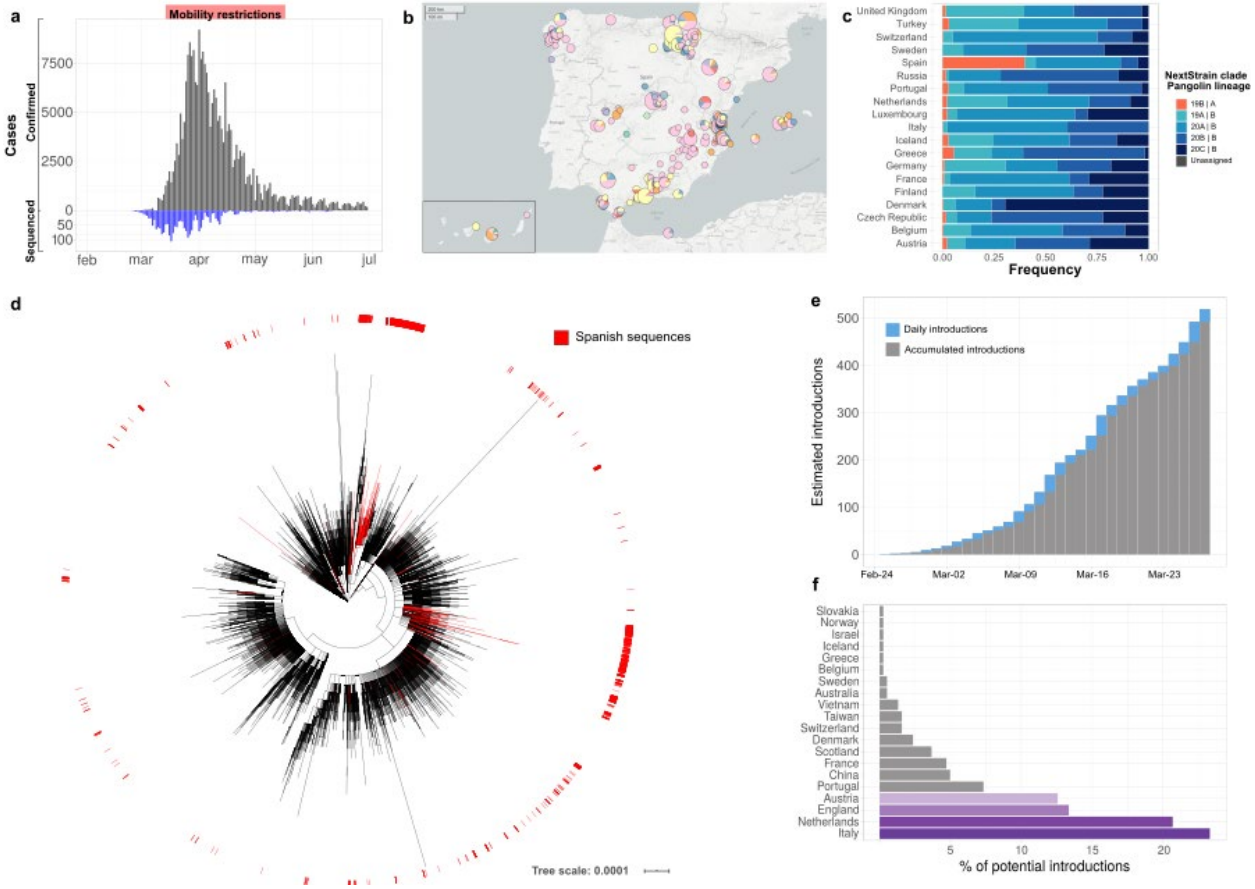


Figure 1. SARS-CoV-2 sequenced genomes from Spain. a. Distribution of sequenced samples (blue) versus confirmed cases in Spain (grey) by date. Country lockdown measures were in effect from 13th March to 17th May 2020. **b.** Distribution of the sequenced samples across Spain was plotted in Microreact. This data can be explored with more detail in the Microreact webpage

(https://microreact.org/showcase) loading the Data S1 files. The size of each piechart correlates with the number of sequences collected in the corresponding area. Each color corresponds to a specific Pangolin lineage, as detailed in Figure S1 (light yellow and green correspond to lineage A, all the others are lineage B). **c.** Distribution of majorSARS-CoV-2 clades during the first stages of the pandemic (before 1st April 2020), in those European countries with more than 50 sequences deposited in GISAID 13th November 2020. **d.** Global maximum likelihood phylogeny constructed with 32,416 sequences, placement of Spanish samples is indicated in red. **e.** New and accumulated introductions to Spain. Lower-bound introduction estimates were defined as the date of the likely infection of the first case in a cluster (14 days before symptom onset). **f.** Estimated international origin of SARS-CoV-2 introductions based on phylogenetic data; in color, those countries with a likely contribution larger than 10%.

A few genetic variants dominated the first wave in Spain

To identify those introductions that resulted in sustained transmission and therefore epidemiologically successful in the long-term, we scanned the phylogeny for larger clades mainly composed by Spanish samples (see Mat and Met for criteria). We identified 9 Spanish Epidemic Clades (SEC) distributed across the phylogeny, representing 46% of the total Spanish dataset analyzed (995 out of 2,170 samples) (Figure 2a, Figure S4, Figure S5, Table S1, Table S2). We first noticed that only two SECs encompassed 30% and 10% of all Spanish samples (SEC8 and SEC7, respectively). This implies that the introduction of these two specific genetic variants explains a high proportion of the entire epidemic for the first wave in the country. In fact, they were responsible for 44% of the 'candidate transmission clusters' identified before the lockdown (Figure 2b). We then estimated the time of introduction in Spain for the 9 SECs using a Bayesian approach (Table S2). As a conservative estimate we considered the time of introduction as any time between the age of the most recent common ancestor of the SEC and the date of the first Spanish sample (Figure 2c). Thus, we assume that the ancestor of the SEC was not necessarily in Spain.

Our analysis shows that the earlier the introduction, the larger the size of the SEC (Figure S6). The larger clades, SEC7 and SEC8, were the first successful genetic variants introduced into Spain during late January - February (Figure 2b). Both belong to lineage A (Pangolin nomenclature) and partially explain the peculiar population structure in Spain relative to other European countries (Figure 1c). In addition, compared with other SECs, SEC7 and SEC8 were widely spread in the country, being present in at least 10 of the 17 administrative regions (Figure 2b) and had a mean pairwise geographic distance between samples of more than 300 km regardless whether or not the Islas Canarias and Baleares are included (Figure S7). On the contrary, SECs that were introduced later were smaller and showed a narrower geographic spread (between 0 - 58 km, ANOVA adjusted p-value << 0.01, Supplementary Notes).

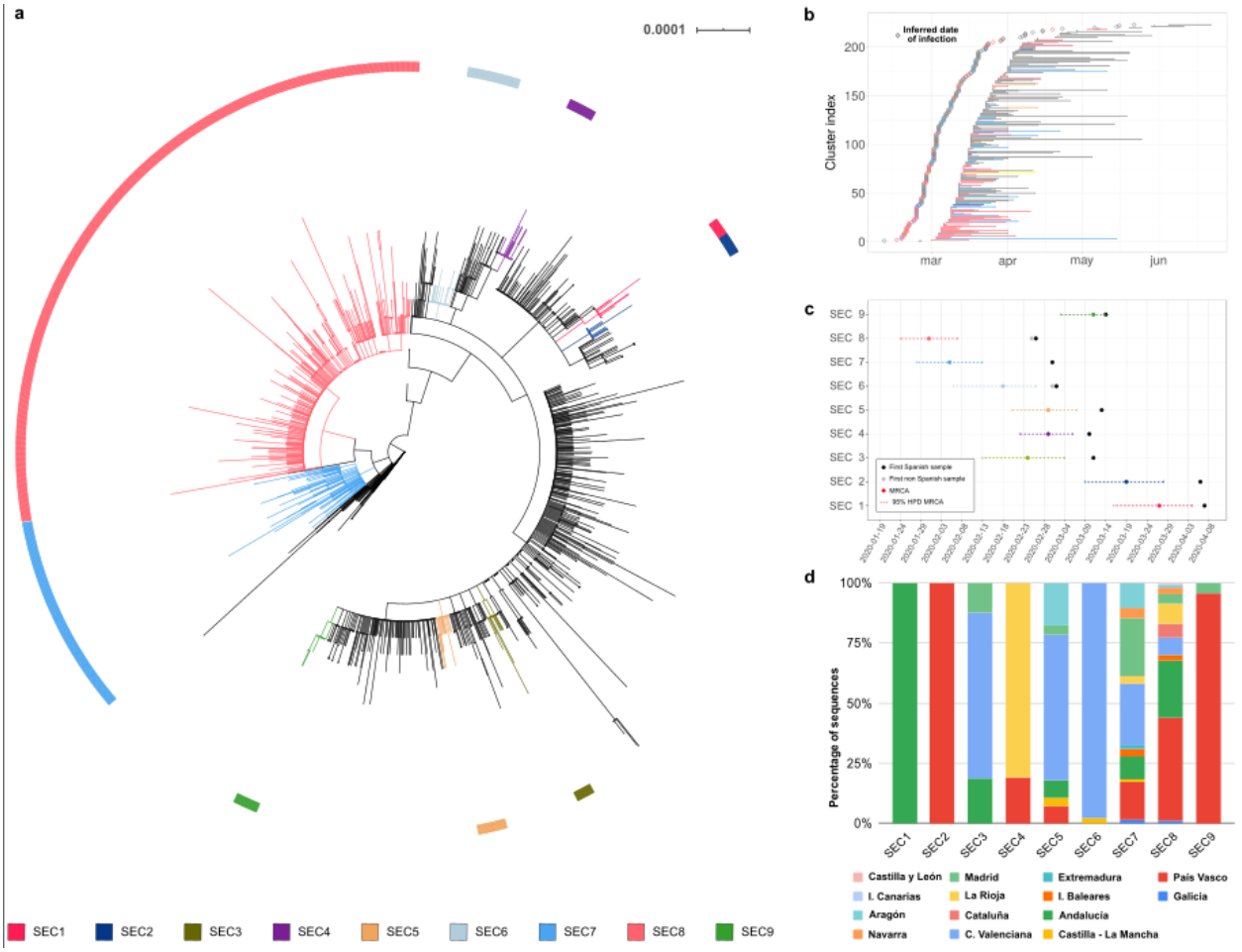


Figure 2. Inferred introduction times and expansion of SECs. a. Maximum likelihood phylogenetic tree of Spanish sequences indicating the identified Spanish Epidemic Clades (SECs). **b.** Range of dates for each 'candidate transmission cluster' identified within the SECs, and the most probable origin date (14 days before the first documented case) **c.** Time of the Most Recent Common Ancestors (MRCA) of each SEC is plotted, including the 95% HPD interval (High Posterior Density). First collected sample is indicated and pointed if it is Spanish or not. **d.** SEC dispersion through the different regions of the country. Some SECs are circumscribed to one or two regions, while some others have expanded through the complete territory.

Superspreading events and mobility were key for the success of SEC8

Why some genetic variants succeed over others cannot be answered solely from genomic sequence data. We must also take into account the epidemiological dynamics in the country. There is data supporting a role of the 614G mutation in the spike protein associated with epidemiological success. However, SEC7 and SEC8 do not harbour the variant, explaining why 614G was less frequent during the first weeks of the epidemic in Spain than in other countries (Figue S8). In addition, the inspection of signature positions for both SECs did not lead to any likely genomic determinant of epidemiological success (Table S3). Alternatively, we have

investigated linked epidemiological data from the earliest cases to shed light on the early success of SEC8.

In a first phase, SEC8 was introduced at least twice from Italy to the city of Valencia (Figure 3a). There is epidemiological evidence that both cases were infected in Italy, as they attended the Atalanta-Valencia Champions League football match on 19th February, and that one of them initiated a transmission chain upon returning to Valencia a few days later. This epidemiological link strongly suggests that the SEC8 genetic variant was imported from Italy. This introduction occurred in agreement with the estimated time of entry of SEC8 into Spain (Table S2). NextStrain tracking tools for viral spatial spread suggests additional SEC8-related early seedings in Madrid, País Vasco, Andalucía, and La Rioja regions (Video S1) which may involve other countries, not necessarily Italy. Given the lack of virus genetic differentiation and scarce epidemiological information there is no certainty on whether they resulted from independent introductions from abroad or from internal migrations of infected persons, although the simultaneous detection in different regions favours the first option. Most of these multiple introductions occurred during the second half of February, a period in which more than 11,000 daily entries of travelers from Italy were recorded.

In a second phase, SEC8 was fueled by superspreading events. Based on the topology of the phylogenetic tree (Figure 2a) there were multiple clades involving a large number of very closely related sequences (1-3 SNPs) (Figure 3a). Of special relevance was a funeral on 23rd February with attendees from the País Vasco and La Rioja regions from which 25 sequences had been sequenced. Importantly, although they did not differ by more than 2 SNPs these sequences are spread across the SEC8 phylogeny suggesting the existence of many more non-sampled secondary cases across the country (Figure 3a). In a third phase, SEC8, after reaching high frequency locally, was redistributed across the country and in less than two weeks it reached a prevalence of 60% among the sequenced genomes (Figure 3b), being present in almost every region analysed. All these phases occurred between the first known diagnosed SEC8 case on 25th February (Table S2) and the lockdown on 14th March, highlighting the need for very early containment measures to stop the spread of SARS-CoV-2.

Effect of lockdown on the major clades

In the second half of March, Spain imposed a strict lockdown on non-essential services and movements. A Bayesian birth-death skyline analysis allowed us to evaluate the impact of the lockdown on the effective reproductive number (R_e) of the most successful SECs. The analyses of SEC7 (Figure S9) and SEC8 (Figure 3c) resulted in similar estimates for R_e before the lockdown (2.10 with 95% highest posterior density, HPD:1.67-2.62 and 3.14 HPD: 2.71-3.57, respectively) similar to the R_e estimated early in the epidemic for SARS-CoV-2[19,20]. After the lockdown there was a substantial decrease to less than 0.5 in both cases (0.27 95% HPD: 0.06-0.47; 0.23 HPD: 0.15-0.32, respectively). The model also estimated that the date with highest support for a change in R_e roughly coincides with the start of the lockdown in Spain on 14th March (20th March HPD: 15-25th March; 9th March HPD: 8-10th March, respectively). In addition, we calculated the doubling time for both SECs[21]. Before the corresponding date of change for R_e, the doubling time for SEC7 was estimated at 6.3 days (95% HPD: 4.3-10.2 days) and that for SEC8 at 3.3 days (95% HPD:

2.7 - 4.1 days). R_e values after those dates had a posterior distribution that did not include 1.0 for both SECs (see Supplementary Notes), a result that supports the reduction in the rate of increase of confirmed cases and that is in agreement with estimates from epidemiological models and data[19,20].

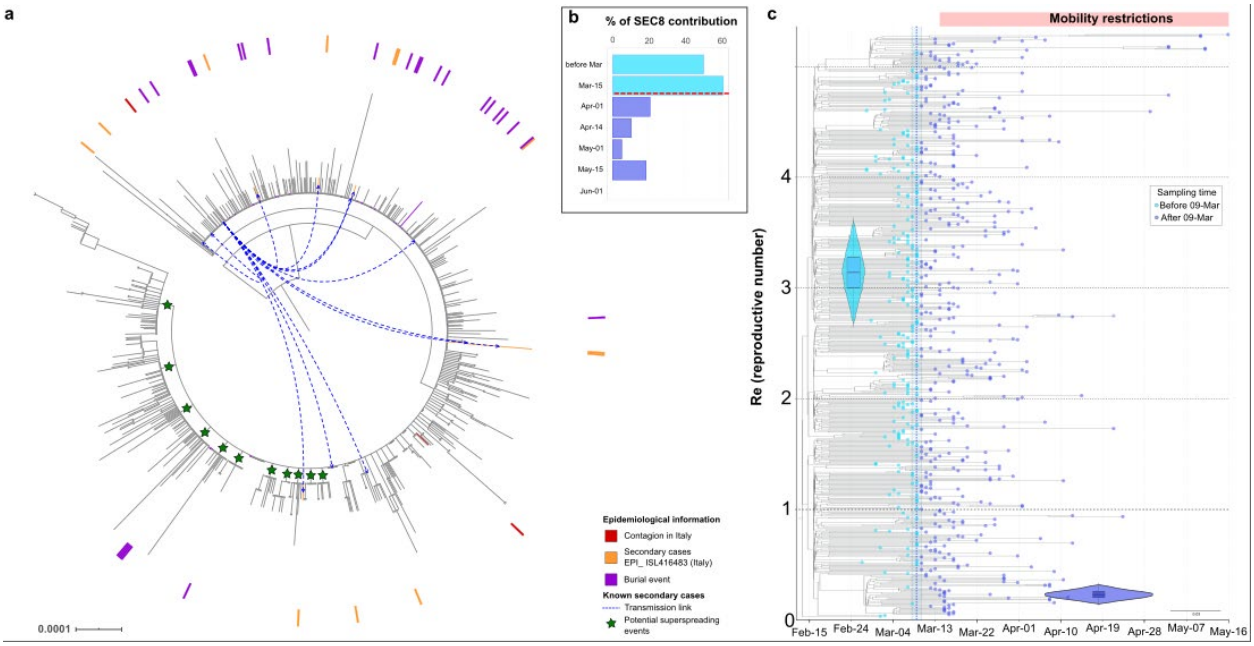


Figure 3. SEC8 epidemiological success and impact of mobility restrictions. a. Maximum likelihood phylogeny with all the strains of SEC8. Samples with epidemiological evidence about their origin are marked in the tree. In red, cases imported from different events in Italy. In orange, secondary cases originated from one of the cases introduced from Italy (also marked with blue arrows). In purple, cases related to a large burial in La Rioja. Green stars mark potential superspreading events of more than 10 sequences sharing at least one clade-defining SNP. **b.** Contribution of SEC8 to the total of samples sequenced over time. The horizontal red line marks the start of the Spanish lockdown, on 14th March. **c.** Phylodynamic estimates of the reproductive number (R_e) of SEC8. The X axis represents time, from the origin of the sampled diversity of SEC8 to the date of the last collected genome on 16th May. The blue dotted line shows the posterior value of the timing of the most significant change in Re, around 9th March [95% HPD: 8–10th March]. The Y axis represents R_e, and the violin plots show the posterior distribution of this parameter before and after the change time in Re, with a mean of 3.14 [95% HPD: 2.71-3.57] and 0.23 [95% HPD: 0.15-0.32] before and after the change time respectively. The phylogenetic tree in the background is a maximum clade credibility tree with the tips colored according to whether they were sampled before or after 9th March.

DISCUSSION

Our analyses have revealed more than 500 independent introductions of SARS-CoV-2 to Spain between late January, coinciding with the first reported cases in our country[2,22], and mid-April 2020. The earliest entries corresponded to lineage A, matching the virus diversity profile reported for the country. This lineage was common in Asia but rare in the rest of Europe[23]. We observed that two genetic variants (SEC7 and SEC8) of this lineage dominated the first stages of the epidemic wave in Spain contrary to what was observed in other European countries. In fact, most cases described in Europe at the beginning were lineage B what makes the situation in Spain more unique. This highlights the importance of epidemiological data in which we know that SEC8 was introduced at least from Italy contradicting the dominant lineages in the country at that time[16,24,25].

Reasons for why some variants dominate over others can be related to the viral genetics, to founder events associated to particular variants, and to the implementation of different measures over time, not necessarily in an exclusive manner. This variant distribution could also be partly explained by sampling bias. No mutation likely associated with epidemiological success has been identified in our analyses of SEC7 and SEC8 (Table S3). In fact, neither SEC7 nor SEC8 carry the 614G mutation in the spike protein contrary to what is seen in most, but not all, lineage B variants (Figure S8). The mutation 614G has been associated with increased viral shedding compared to the ancestral 614D variant in laboratory conditions[26] and in transmission studies[27,28]. However, other studies cast doubts on its actual role in the epidemic[29] suggesting that its impact on epidemic transmission was minor, if any. In the case of Spain, 614G was not behind the initial success of the epidemic because SEC7 and particularly SEC8 were much more common than other genetic variants until the lockdown (10% and 30% of cases respectively). On the contrary, founder events seem to have played an important role for these particular variants. Our analysis shows that they were the first variants introduced in the country and, at least SEC8, were linked to very early superspreading events that contributed to their success. However, an early introduction of lineage A variants also occurred in other European countries but they did not take hold and were displaced by lineage B. Despite the early adoption of strong NPI measures, we hypothesize that epidemic control in the first wave in Spain was soon overwhelmed as compared to countries that controlled early outbreaks[13]. This was likely associated with a strict implementation of the case definition by the WHO, which allowed a stealth dispersion of the first introductions, but also to several superspreading events, which strongly favoured the establishment of the earliest variants arriving into the country. Spain implemented one of the most strict lockdowns in Europe with a high compliance from the population as tracked by mobility data[30]. The efficacy of NPI measures was evident a few weeks later and it was reflected in the almost complete elimination of SEC7 and SEC8 by the end of the first wave. The spread of new variants, represented by other SECs and more isolated cases, corresponded to a new phase in the epidemic at the national level, with much more limited mobility and social interactions which prevented the establishment of large clusters and transmission chains except in high risk settings such as nursing-homes and long-term care facilities.

This study has several limitations. Despite being one of the countries with more contribution to public repositories, our dataset only represents a small subset of confirmed cases that occurred in the first COVID-19 wave (1% of cases). Moreover, sampling across the country was

heterogeneous and the representation of each region in the dataset was not always proportional to the incidence during the studied period. Lack of genome data from countries with high disease burden, especially at the beginning of the pandemic, may have led to underestimating the total number of introductions and prevented a reliable identification of their likely sources based only on viral genome sequences. In addition, we did not have access to individual patient data for most cases. Despite these limitations, we have been able to investigate some of the key cases and events that ignited the epidemic in Spain. This allowed us to understand the origin and early spread of SEC8, which would not have been possible based only on genome data. But we have also shown that genetic data can be used to accurately estimate relevant epidemiological parameters such as R_e and doubling times even when the proportion of sampling is low.

We believe that our results allow us to draw lessons for the control of this and future pandemics. First, we have shown how specific variants can be used to track the effectiveness of epidemic control measures. In February, the number of SEC8 cases was just a few dozens and yet it ended up accounting for 60% of the sequenced samples in the first weeks of March. Second, the closure of borders to countries with high incidence is relevant to reduce simultaneous and multiple imports of the virus, but its efficacy depends on the inward incidence of the disease[31]. The most successful SECs during the first wave were probably those that arrived early, multiple times, and to diverse locations. Thus, as suggested elsewhere, founder effects are important for the success of certain variants. Third, SEC7 and SEC8 extended across Spain in a matter of days. Controlling mobility is essential when the level of community transmission is high, as demonstrated by the significant decrease in R_e for these high-transmission genetic variants after the lockdown. As a comparison, before the lockdown R_e values were 50% higher in Spain (3.3 for SEC8) than in Australia (1.63), and they underwent a reduction down to 7% of the original value (0.23) as a result of the containment measures, compared to 30% (0.48) in Australia[15]. From a public health perspective, our results add to the evidence that the success of specific genetic variants is fueled by superspreading events which rapidly increase the prevalence of the virus[32]. Subsequently, coupled to the high mobility of our connected world, a variant may end up dominating the epidemic in a geographic location. This is what occurred to SEC8 and what at a local level has been described in Boston[33]. In fact, we have recently described a new variant in Europe that is rapidly growing in several countries, which is also linked to initial superspreading events[34]. The conclusion is that early diagnosis and notification of cases would have helped to a timely implementation of effective contact tracing that, coupled with earlier mobility closures and maybe tighter border control, would have probably delayed a few days the expansion of genetic variants such SEC8 during the early stages of the epidemic in Spain. Whether this might have changed the global shape of the epidemic in the country or other genetic variants would have performed its role leading to a similar outcome cannot be ascertained, but the comparison with other countries lead us to suspect that there would have been not many differences with them.

REFERENCES

1. Zhu, N. *et al.* Brief report: A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727 (2020).

2. Spiteri, G. *et al.* First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. *Eurosurveillance* **25**, (2020).
3. Pollán, M. *et al.* Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
4. Ministerio de la Presidencia, R. C. las C. y. M. D. BOE-A-2020-3692. *https://www.boe.es/eli/es/rd/2020/03/14/463/con* (2020).
5. Centro de Coordinación de Alertas y Emergencias Sanitarias, Ministerio de Sanidad. *Enfermedad por el coronavirus (COVID-19)*. https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_268_COVID-19.pdf (2020).
6. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004).
7. Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J. & Frost, S. D. W. Phylodynamics of Infectious Disease Epidemics. *Genetics* **183**, 1421–1430 (2009).
8. Vasylyeva, T. I. *et al.* Phylodynamics Helps to Evaluate the Impact of an HIV Prevention Intervention. *Viruses* **12**, (2020).
9. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, (2017).
10. Callaway, E. The coronavirus is mutating — does it matter? *Nature* vol. 585 174–177 (2020).
11. Kupferschmidt, K. The pandemic virus is slowly mutating. But does it matter? *Science* **369**, 238–239 (2020).
12. Morel, B. *et al.* Phylogenetic analysis of SARS-CoV-2 data is difficult. *bioRxiv* 2020.08.05.239046 (2020) doi:10.1101/2020.08.05.239046.
13. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).

14. Geoghegan, J. L. *et al.* Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *medRxiv* (2020) doi:10.1101/2020.08.05.20168930.
15. Seemann, T. *et al.* Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **11**, 4376 (2020).
16. Alm, E. *et al.* Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Eurosurveillance* **25**, 2001410 (2020).
17. Oude Munnink, B. B. *et al.* Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **26**, 1405–1410 (2020).
18. Candido, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**, 1255–1260 (2020).
19. Guirao, A. The Covid-19 outbreak in Spain. A simple dynamics model, some lessons, and a theoretical framework for control response. *Infectious Disease Modelling* **5**, 652 (2020).
20. Hyafil, A. & Moriña, D. Analysis of the impact of lockdown on the reproduction number of the SARS-Cov-2 in Spain. *Gac. Sanit.* (2020) doi:10.1016/j.gaceta.2020.05.003.
21. Lurie, M. N., Silva, J., Yorlets, R. R., Tao, J. & Chan, P. A. Coronavirus Disease 2019 epidemic doubling time in the United States before and during stay-at-home restrictions. *J. Infect. Dis.* **222**, 1601–1606 (2020).
22. Böhmer, M. M. *et al.* Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases* **20**, 920–928 (2020).
23. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* **5**, 1403–1407 (2020).
24. Lai, A. *et al.* Molecular Tracing of SARS-CoV-2 in Italy in the First Three Months of the Epidemic. *Viruses* **12**, (2020).

25. Phylogenetic analysis of SARS-CoV-2 diversity in Europe (Italy). https://nextstrain.org/groups/neherlab/ncov/italy.
26. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 1–6 (2020).
27. Volz, E. M. *et al.* Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. *Cell* (2020) doi:10.1016/j.cell.2020.11.020.
28. Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827 (2020).
29. van Dorp, L. *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat. Commun.* **11**, 1–8 (2020).
30. Google. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/.
31. Russell, T. W. *et al.* Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *The Lancet Public Health* (2020) doi:10.1016/s2468-2667(20)30263-2.
32. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* (2020) doi:10.1126/scitranslmed.abe2555.
33. Lemieux, J. *et al.* Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and superspreading events. *medRxiv* (2020) doi:10.1101/2020.08.23.20178236.
34. Hodcroft, E. B. *et al.* Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* 2020.10.25.20219063 (2020).

SUPPLEMENTARY MATERIAL

- Supplementary Notes
- Supplementary Material and Methods
- Supplementary References

- Supplementary Data
  - Table S1: GISAID accession numbers for the 32914 sequences used in this study. The 'basal group' used for dating and the sequences representative of the pangolin lineages are marked for identification.
  - Table S2: SEC characteristics and inferred origin time. The time of the most recent common ancestors (MRCA) of each SEC was estimated with a Bayesian molecular clock analysis. "MRCA date" indicates the median value for the age of the closest SEC MRCA. The 95% Highest Posterior Density (HPD) credibility interval for this value is provided. "SEC size" indicates the number of samples belonging to each SEC. The first Spanish collected sample within each SEC is also indicated; the inferred date of infection is inferred as the time span between the oldest MRCA date and the first Spanish collected sample. Number of "candidate transmission clusters", "zero distance clusters" and "unique" included in each SEC are mentioned. "MRCA2" indicates the time of the previous ancestor to the MRCA, considering only nodes that display a posterior probability higher than 0.5. If we consider that the MRCAs were already in Spain, then the introductions into the country occurred between the MRCA2 and the MRCA dates.
  - Table S3: SEC definitory mutations.
  - Data S1.zip: Alignment and phylogenetic tree of the Spanish sequenced samples, to be plotted using the Microreact server.
  - Video S1: Video showing the transmission dynamics of SEC8 within Spain from 2020-01-16 to 2020-07-19, obtained from NextStrain.
- Supplementary Figures

-Figure S1: Abundance of the different Pangolin lineages in the dataset by epidemiological week (number of weeks since 2019-12-23) as plotted in Microreact.

-Figure S2: Examples of the different groups of sequences identified. 'Candidate transmission clusters' are groups of Spanish sequences that form a clade. 'Zero distance clusters' are groups of Spanish sequences that are at zero distance from each other. Finally, the 'unique' sequences are Spanish sequences that are more than 1 SNP away from any other Spanish sequence and that do not share a most recent common ancestor (MRCA) node with other Spanish sequences

-Figure S3: Distribution of the different clusters/groups sizes in Spanish samples.

-Figure S4: Number of international and Spanish sequences in each SEC.

-Figure S5: Phylogenetic location of each SEC in the global SARS-CoV-2 phylogeny. Sequences from Spain are coloured according to their SEC (as indicated in Figure 2) while international sequences remain in black colour.

-Figure S6: Time of the MRCA of each SEC plotted against the contribution of each SEC to the total number of samples in the Spanish dataset. We observed a significant correlation (ρ=-0.69, p-value=0.03) between the time of the MRCA of each SEC and its size, estimated as the number of samples sequenced.

-Figure S7: Distribution of genetic (salmon) versus geographic (grey) distances within each pair of samples belonging to the same SEC.

-Figure S8: Distribution of sequences harbouring the 614G mutation (blue) versus the 614D mutation (salmon,wild-type) in the S gene for the spanish sequences in our dataset. In the

left panel, a histogram of samples sorted by date of sequencing. At right, frequency of both mutations in the sequenced samples by date. The national lockdown event is marked by a purple vertical line.

-Figure S9: Phylodynamic estimates of the effective reproductive number (R_e) of Spanish SEC7. A birth–death skyline (BDSKY) model was implemented in Beast v.2, allowing for piecewise changes in Re, with the time and magnitude estimated from the data. The X axis represents time, starting with the MRCA of all sampled diversity within SEC7 and ending with the date of the most recently sequenced genome from 15th May. The blue dotted line indicates the posterior value of the timing of a most significant decrease in Re, around 20th March [95% HPD: 15–25th March]. The Y axis represents Re, and the violin plots show the posterior probability distribution for this parameter before and after the change time in Re; with a mean of 2.10 [95% HPD: 1.67–2.62] and 0.27 [95% HPD:0.06–0.47] before and after this time, respectively. The phylogenetic tree in the background is the maximum clade credibility tree from the BDSKY analysis, with the tips colored according to whether they were sampled before or after 20th March.

- Figure S10: Mean pairwise genetic distance vs geographic distance (in SNP number), between the largest cities (> 70k inhabitants) of the Comunidad Valenciana autonomous region.

- Figure S11: Left) Heatmap of genetic diversity for the province of Valencia; red colors indicate high diversity; blue colors indicate lower diversity. Genetic diversity has been measured as the number of base substitutions per site averaged over all sequence pairs within each municipality. Genetic diversity is largest near Valencia, the region's capital, and decreases with geographic distance to it. Right) All sequences included in our dataset from Comunidad Valenciana, coloured according to the pangolin lineage they belong to.

Acknowledgements

This work was funded by the Instituto de Salud Carlos III project COV20/00140, Spanish National Research Council project CSIC-COV19-021 and ERC StG 638553 to IC, and BFU2017-89594R to FGC. MC is supported by Ramón y Cajal program from Ministerio de Ciencia and grants RTI2018-094399-A-I00 and SEJI/2019/011.

We gratefully acknowledge Hospital Universitari Vall d'Hebron, Instituto de Salud Carlos III, IrsiCaixa AIDS Research Lab and all the international researchers and institutions that submitted sequenced SARS-CoV-2 genomes to the GISAID's EpiCov™ Database, as an important part of our analyses have been made possible by the share of their work.

Author contributions

IC, FGC and MC conceived the work. GAG, GDA and SJS set up the bioinformatics environment and the analysis pipeline. MGL, ACO, PRR and NGG analysed the data. ACO and MGL wrote the first version of the draft. AO, JR, EM, AEB, AN, DGV, LPL, MH, JS, MAM, MT, MPBE, NGJ, GM, LMP, PRH, LRR, MTP, IGN, JFP sequenced genomes. GCE, MMR, LPV, JMM, RMM,

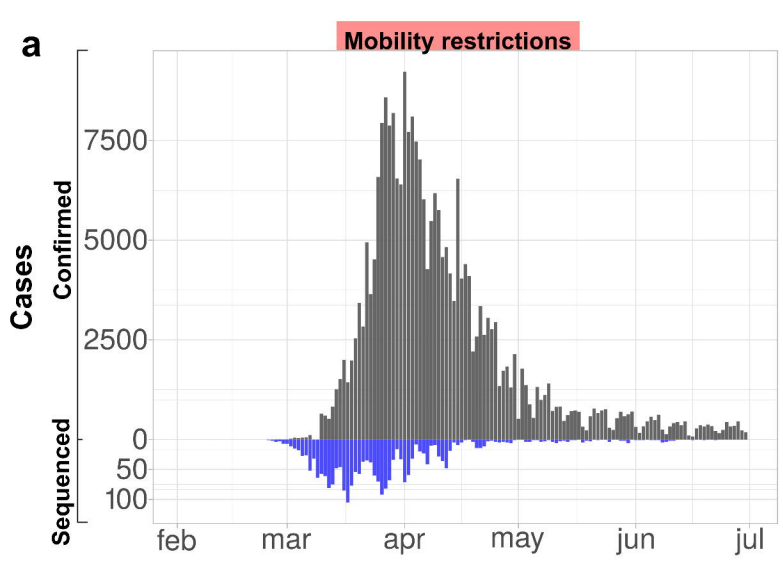
MDTB, JAL, VGG, ARV, DN, EA, IT, ACB, SHC, MLCR, AR, AT, AMB, JMNM, IPC, SSG, BFE, CGC, BPB, ITP, AC, VM, MPG, LRF, JLP, JA, JJCA, MCPG, JAB, NR, JLLH, MAZ provided samples. IC, FGC, MC, ACO, MGL, SD and DGV critically reviewed and contributed to the final version of the paper.

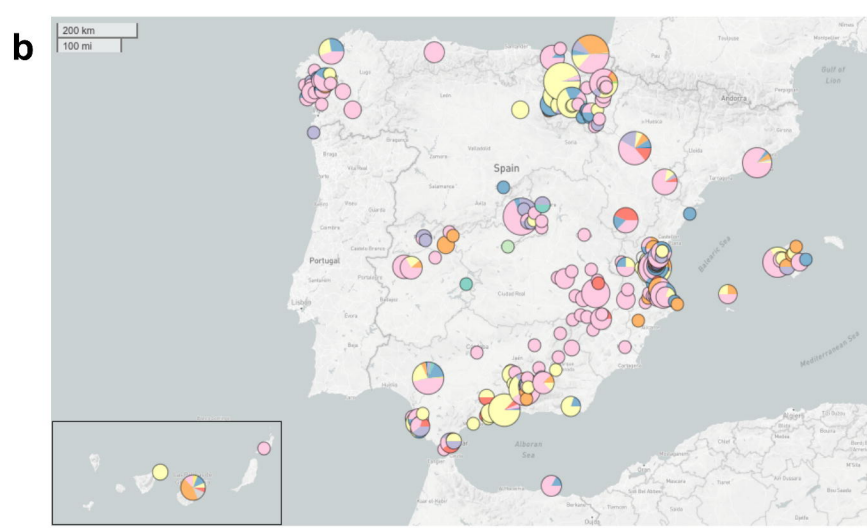
Annex I - List of the SeqCOVID-SPAIN consortium members

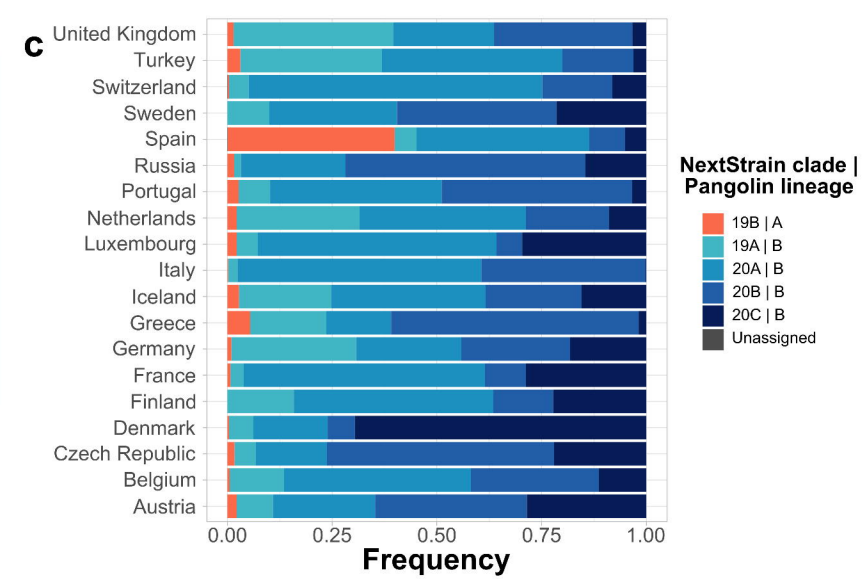
Iñaki Comas (icomas@ibv.csic.es), Fernando González-Candelas (fernando.gonzalez@uv.es), Galo Adrian Goig (galo.goig@swisstph.ch), Álvaro Chiner-Oms (achiner@ibv.csic.es), Irving Cancino-Muñoz (icancino@ibv.csic.es), Mariana Gabriela López (mglopez@ibv.csic.es), Manoli Torres-Puente (mtorres@ibv.csic.es), Inmaculada Gómez-Navarro (igomez@ibv.csic.es), Santiago Jiménez-Serrano (sjimenez@ibv.csic.es), Lidia Ruiz-Roldán (lidiarroldan@gmail.com), María Alma Bracho (bracho_alm@gva.es), Neris García-González (neris@uv.es), Llúcia Martínez-Priego (martinez_lucpri@gva.es), Inmaculada Galán-Vendrell (galan_inm@gva.es), Paula Ruiz-Hueso (ruiz_pau@gva.es), Griselda De Marco (demarco_gri@gva.es), Mª Loreto Ferrús (ferrus_mlo@gva.es), Sandra Carbó-Ramírez (carbo_sanram@gva.es), Mireia Coscollá (mireia.coscolla@uv.es), Paula Ruiz-Rodríguez (ruizro5@alumni.uv.es), Giuseppe D'Auria (dauria_giu@gva.es), Francisco Javier Roig Sena (roig_fco@gva.es), Isabel Sanmartín (isanmartin@rjb.csic.es), Daniel García-Souto (danielgarciasouto@gmail.com), Ana Pequeño-Valtierra (ana.pequeno@usc.es), Jose M. C. Tubio (jmctubio@gmail.com), Javier Temes (fjtemes@gmail.com), Jorge Rodríguez-Castro (jorge.rodriguez@usc.es), Martín Santamarina (martin.santamarina.garcia@usc.es), Nuria Rabella (nrabella@santpau.cat), Ferrán Navarro (fnavarror@santpau.cat), Elisenda Miró (emiro@santpau.cat), Manuel Rodríguez-Iglesias (manuel.rodrigueziglesias@uca.es), Fátima Galán-Sanchez (fatima.galan@uca.es), Salud Rodríguez-Pallares (salud361@gmail.com), María de Toro (mthernando@riojasalud.es), María Pilar Bea-Escudero (mpbea@riojasalud.es), José Manuel Azcona-Gutiérrez (jmazcona@riojasalud.es), Miriam Blasco-Alberdi (mblasco@riojasalud.es), Alfredo Mayor (alfredo.mayor@isglobal.org), Alberto L. García-Basteiro (alberto.garcia-basteiro@isglobal.org), Gemma Moncunill (gemma.moncunill@isglobal.org), Carlota Dobaño (carlota.dobano@isglobal.org), Pau Cisteró (pau.cistero@isglobal.org), Darío García de Viedma (dgviedma2@gmail.com), Laura Pérez-Lago (lperezg00@gmail.com), Marta Herranz (m_herranz01@hotmail.com), Jon Sicilia (jsiciliamambrilla@gmail.com), Pilar Catalán (pcatalan.hgugm@salud.madrid.org), Julia Suárez (julia.suarez@iisgm.com), Patricia Muñoz (pmunoz@hggm.es), Cristina Muñoz-Cuevas (cristina.munozc@salud-juntaex.es), Guadalupe Rodríguez Rodríguez (guadalupe.rodriguez@salud-juntaex.es), Juan Alberola (juan.alberola@uv.es), Jose Miguel Nogueira Coito (Jose.M.Nogueira@uv.es), Juan José Camarena Miñana (juan.camarena@uv.es), Antonio Rezusta (arezusta@unizar.es), Alexander Tristancho (aitristancho@salud.aragon.es), Ana Milagro Beamonte (amilagro@salud.aragon.es), Nieves Martínez Cameo (nmcameo@gmail.com), Yolanda Gracia Grataloup (ygrataloup@yahoo.es), Elisa Martró (emartro@igtp.cat), Antoni E. Bordoy (aescalas@igtp.cat), Anna Not (anot@igtp.cat), Adrián Antuori (adrian.antuori@gmail.com), Anabel Fernández (afernandezn.ics@gencat.cat), Nona Romaní (nonaromani@gmail.com), Rafael Benito (rbenito@unizar.es), Sonia Algarate Cajo (sonialgarate@gmail.com), Jessica

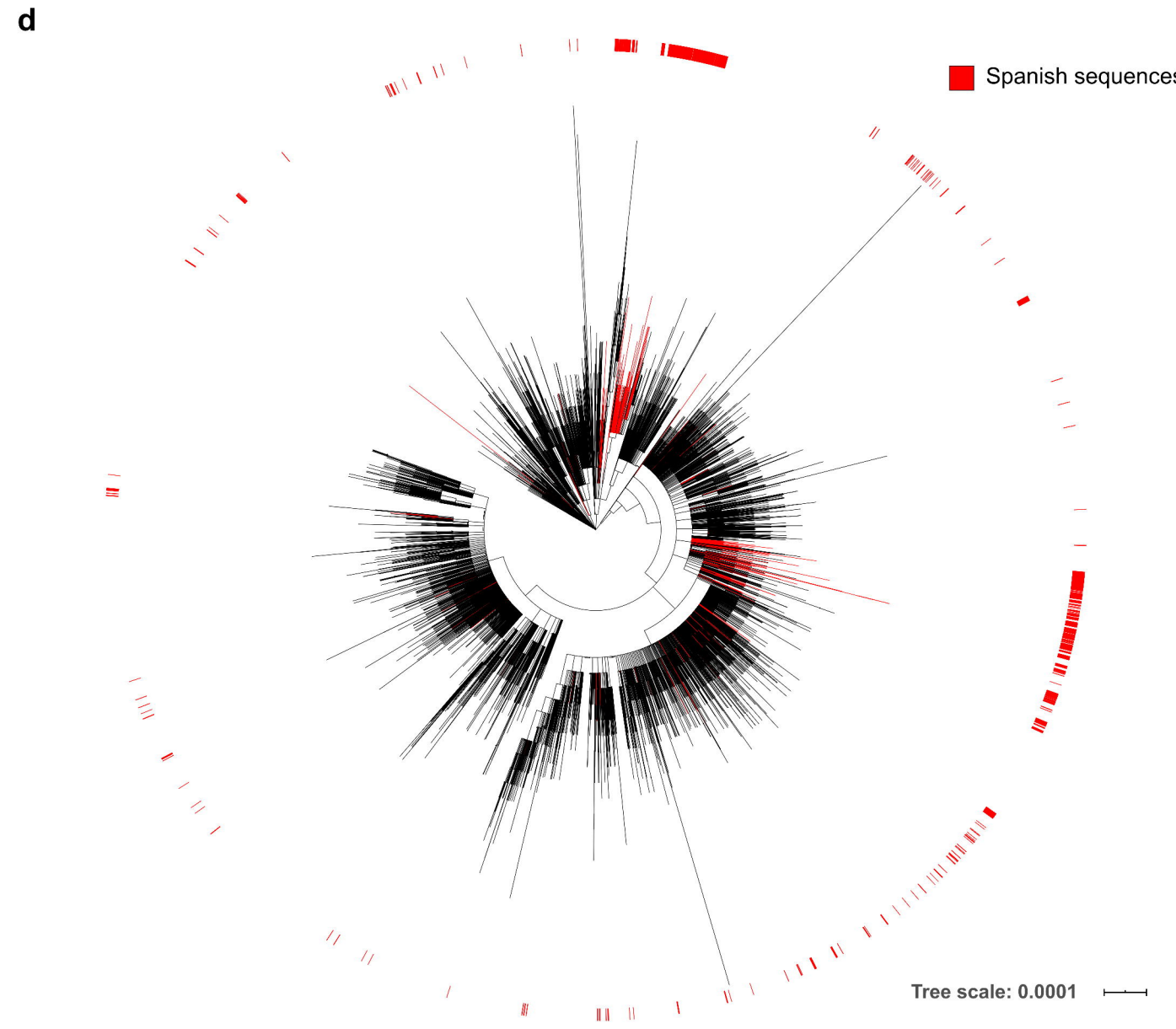
Bueno (jbuenosan@salud.aragon.es), Jose Luis del Pozo (jdelpozo@unav.es), Jose Antonio Boga (joseantonio.boga@sespa.es), Cristián Castelló Abietar (crcaab@hotmail.com), Susana Rojo Alba (ssnrj4@gmail.com), Marta Elena Álvarez Argüelles (martaealvarez@gmail.com), Santiago Melón García (santiago.melon@sespa.es), Maitane Aranzamendi Zaldumbide (maitane.aranzamendizaldumbide@osakidetza.eus), Andrea Vergara (vergara@clinic.cat), Miguel J Martínez (myoldi@clinic.cat), Jordi Vila (jvila@clinic.cat), David Posada (dposada@uvigo.es), Diana Valverde (dianaval@uvigo.es), Nuria Estévez (nuestevez@uvigo.es), Iria Fernández-Silva (irfernandez@uvigo.es), Loretta de Chiara (Ldechiara@uvigo.es), Pilar Gallego (mgallego@alumnos.uvigo.es), Nair Varela (nvarela@alumnos.uvigo.es), Rosario Moreno-Muñoz (moreno_rosmuny@gva.es), Mª Dolores Tirado Balaguer (tirado_dolbal@gva.es), Ulises Gómez-Pinedo (ulisesalfonso.gomez@salud.madrid.org), Mónica Gozalo Margüello (monica.gozalo@scsalud.es), Mª Eliecer Cano García (meliecer.cano@scsalud.es), José Manuel Méndez Legaza (josemanuel.mendez@scsalud.es), Jesús Rodríguez Lozano (jesus.rodriguez@scsalud.es), María Siller Ruiz (maria.siller@scsalud.es), Daniel Pablo Marcos (daniel.pablo@scsalud.es), Antonio Oliver (aoliverp@yahoo.es), Jordi Reina (jordireina1957@gmail.com), Carla López-Causapé (carla.lopez393@gmail.com), Andrés Canut-Blasco (andres.canutblasco@osakidetza.eus), Silvia Hernáez Crespo (silvia.hernaezcrespo@osakidetza.eus), María Luz Cordón Rodríguez (marialuzalbina.cordonrodriguez@osakidetza.eus), Mª Concepción Lecaroz Agara (mariaconcepcion.lecarozagara@osakidetza.eus), Carmen Gómez González (carmen.gomezgonzalez@osakidetza.eus), Amaia Aguirre Quiñonero (amaia.aguirrequiñonero@osakidetza.eus), José Israel López Mirones (joseisrael.lopezmirones@osakidetza.eus), Marina Fernández Torres (marina.fernandeztorres@osakidetza.eus), Mª Rosario Almela Ferrer (mariadelrosario.almelaferrer@osakidetza.eus), José Antonio Lepe (josea.lepe.sspa@juntadeandalucia.es), Verónica González Galán (veronica.gonzalez.galan.sspa@juntadeandalucia.es), Ángel Rodríguez-Villodres (angel.rodriguez.villodres.sspa@juntadeandalucia.es), Nieves Gonzalo Jiménez (gonzalo_nie@gva.es), Maria Montserrat Ruiz García (ruiz_mongar@gva.es), Antonio Galiana (antoniogaliana1@gmail.com), Judith Sánchez-Almendro (judhsa@gmail.com), Gustavo Cilla Eguiluz (carlosgustavosantiago.cillaeguiluz@osakidetza.eus), Milagrosa Montes Ros (mariamilagrosa.montesros@osakidetza.eus), Luis Piñeiro Vázquez (luisdario.pineirovazquez@osakidetza.eus), Ane Sorrarain (ane.sorarrain@biodonostia.org), José María Marimón (josemaria.marimonortizdez@osakidetza.eus), Maria Dolores Gómez Ruiz (gomez.mdo@gmail.com), Eva M. González-Barberá (gonzalez_evabar@gva.es), José Luis López-Hontangas (lopez_jlu@gva.es), José María Navarro-Marí (josem.navarro.sspa@juntadeandalucia.es), Irene Pedrosa-Corral (irene.pedrosa.sspa@juntadeandalucia.es), Sara Luisa Sanbonmatsu-Gámez (saral.sanbonmatsu.sspa@juntadeandalucia.es), Maria Carmen Perez Gonzalez (mcpergon@gobiernodecanarias.org), Francisco Javier Chamizo López (fchalop@gobiernodecanarias.org), Ana Bordes Benítez (aborben@gobiernodecanarias.org), David Navarro (david.navarro@uv.es), Eliseo Albert (eliseo.al.vi@gmail.com), Ignacio Torres (nachotfink@gmail.com), María Isabel Gascón Ros (gascon_isa@gva.es), Cristina Torregrosa

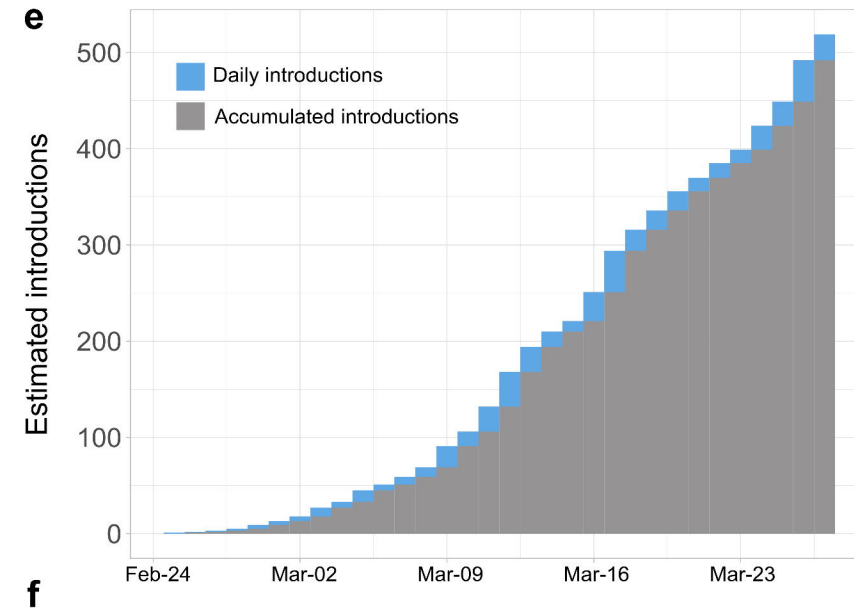
694 Hetland (cjtorregrosahetland@hotmail.com), Eva Pastor Boix (pastor_eva@gva.es), Paloma
695 Cascales Ramos (cascales_pal@gva.es), Begoña Fuster Escrivá (begona.fuster@gmail.com),
696 Concepción Gimeno Cardona (concepcion.gimeno@uv.es), María Dolores Ocete Mochón
697 (ocete_mar@gva.es), Rafael Medina González (rafa.medina.gonzalez@gmail.com), Julia
698 González-Cantó (juliagonzalez1992@hotmail.com), Olalla Martínez (martinez_ola@gva.es),
699 Begoña Palop-Borrás (bpalop@hotmail.com), Inmaculada de Toro Peinado
700 (inmadetoro@yahoo.es), María Concepción Mediavilla Gradolph (gradolphilla@hotmail.com),
701 Oscar González-Recio (gonzalez.oscar@inia.es), Mónica Gutiérrez-Rivas
702 (mgrivas9@gmail.com), Encarnación Simarro Córdoba (mesimarro@sescam.jccm.es), Julia
703 Lozano Serra (jlozanos@sescam.jccm.es), Mónica Parra Grande
704 (monicaparra88@hotmail.com), Lorena Robles Fonseca (lrobles@sescam.jccm.es), Adolfo de
705 Salazar (adolsalazar@gmail.com), Laura Viñuela (lauravinuelagon@gmail.com), Natalia
706 Chueca (naisses@yahoo.es), Federico García (fegarcia@ugr.es), Cristina Gomez-Camarasa
707 (gomezcamarasa@gmail.com), Ana Carvajal (ana.carvajal@unileon.es), Vicente Martín
708 (vicente.martin@unileon.es), Juan Miguel Fregeneda-Grandes (juan.fregeneda@unileon.es),
709 Antonio José Molina (ajmolt@unileon.es), Héctor Argüello (hector.arguello@unileon.es), Tania
710 Fernandez-Villa (tferv@unileon.es), Amparo Farga Martí (farga_amp@gva.es), Rocío Falcón
711 (falcon_roc@gva.es), Victoria Domínguez Márquez (m.victoria.dominguez@uv.es), José Javier
712 Costa-Alcalde (jose.javier.costa.alcalde@sergas.es), Rocío Trastoy
713 (rocio.trastoy.pena@sergas.es), Gema Barbeito-Castiñeiras
714 (gema.barbeito.castineiras@sergas.es), Amparo Coira (amparo.coira.nieto@sergas.es), María
715 Luisa Pérez del Molino Bernal (maria.luisa.perez.del.molino.bernal@sergas.es), Antonio
716 Aguilera (antonio.aguilera.guirao@sergas.es), Anna Planas (anna.planas@iibb.csic.es), Álex
717 Soriano (asoriano@clinic.cat), Israel Fernández-Cádenas (israelcadenas@yahoo.es), Jordi
718 Pérez-Tur (jpereztur@ibv.csic.es), María Ángeles Marcos (mmarcos@clinic.cat), Manuel
719 Segovia Hernández (msegovia@um.es), Antonio Moreno Docón (a.moreno@um.es), Juan
720 Carlos Galan (juancarlos.galan@salud.madrid.org), Esther Viedma Moreno
721 (viedmaesther@gmail.com), Jesús Mingorance (jesus.mingorance@idipaz.es), Jovita
722 Fernández-Pinero (fpinero@inia.es), Elisa Rubio García (elrubio@clinic.cat), Aida Peiró-
723 Mestres (aida.peiro@isglobal.org), Jessica Navero-Castillejos
724 (jessica.navero@isglobal.org)

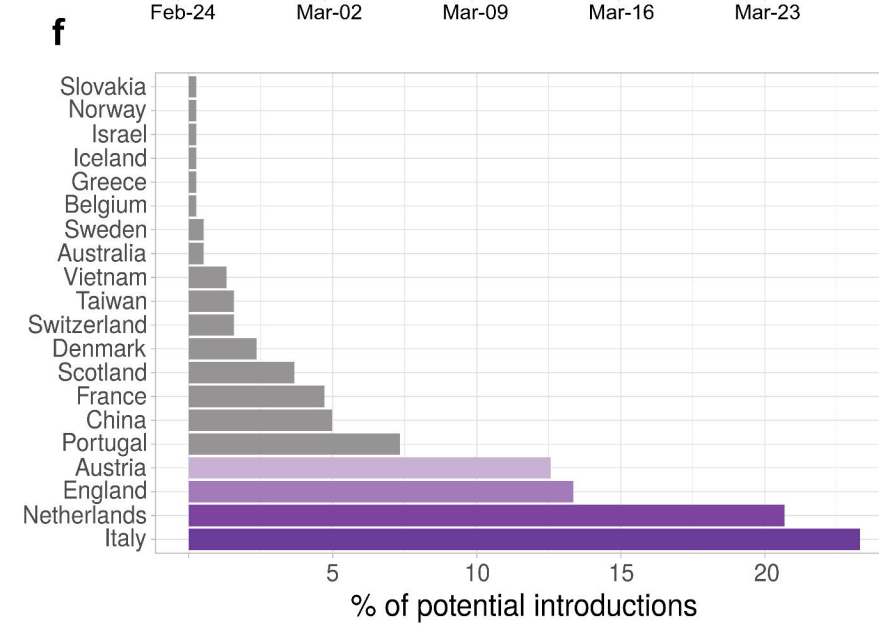

a
Mobility restrictions
Cases
Confirmed
Sequenced
7500
5000
2500
0
50
100
feb
mar
apr
may
jun
jul


b
200 km
100 mi
Spain
Portugal

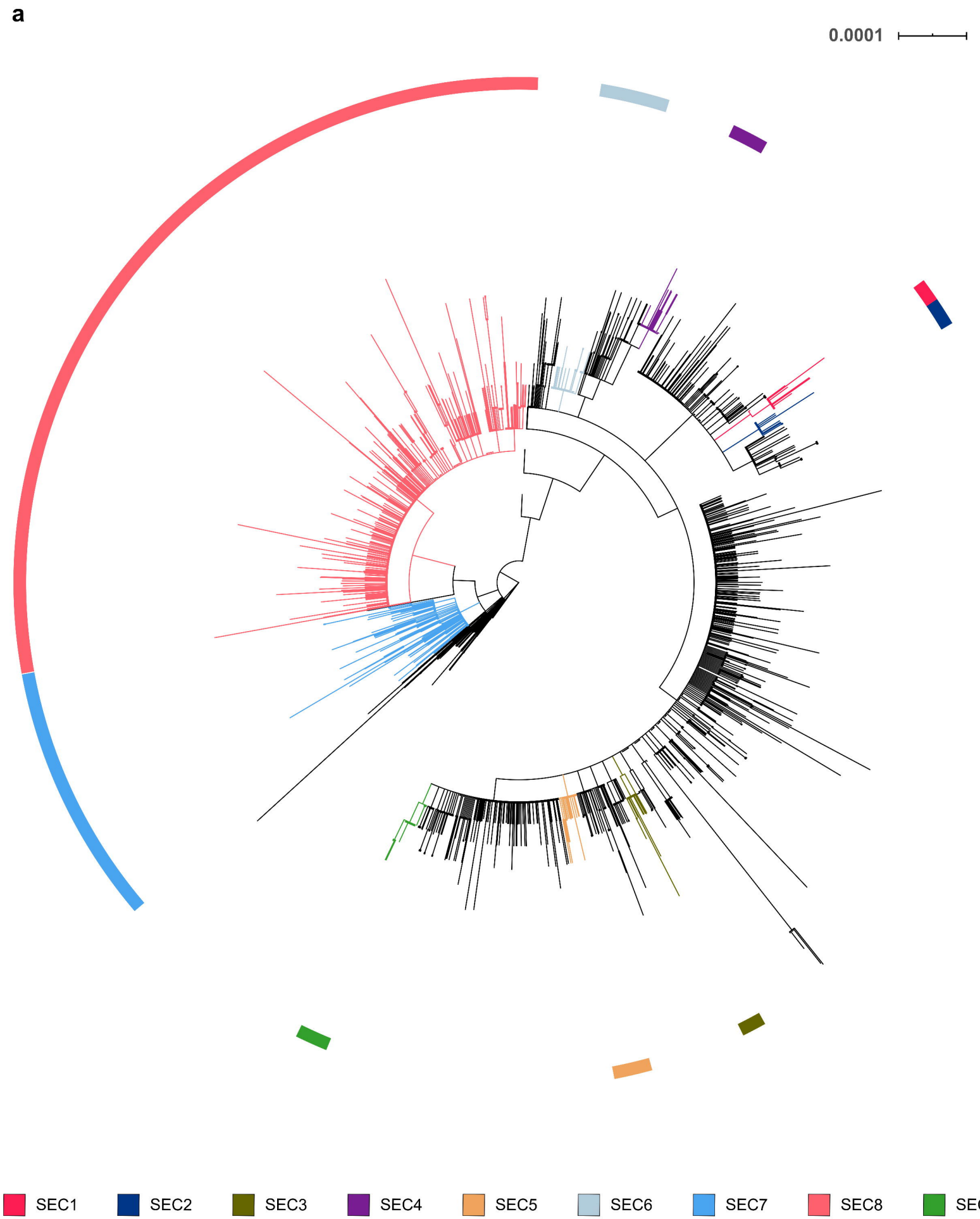

c
United Kingdom
Turkey
Switzerland
Sweden
Spain
Russia
Portugal
Netherlands
Luxembourg
Italy
Iceland
Greece
Germany
France
Finland
Denmark
Czech Republic
Belgium
Austria
0.00
0.25
0.50
0.75
1.00
Frequency
NextStrain clade | Pangolin lineage
19B | A
19A | B
20A | B
20B | B
20C | B
Unassigned


d
Spanish sequences
Tree scale: 0.0001

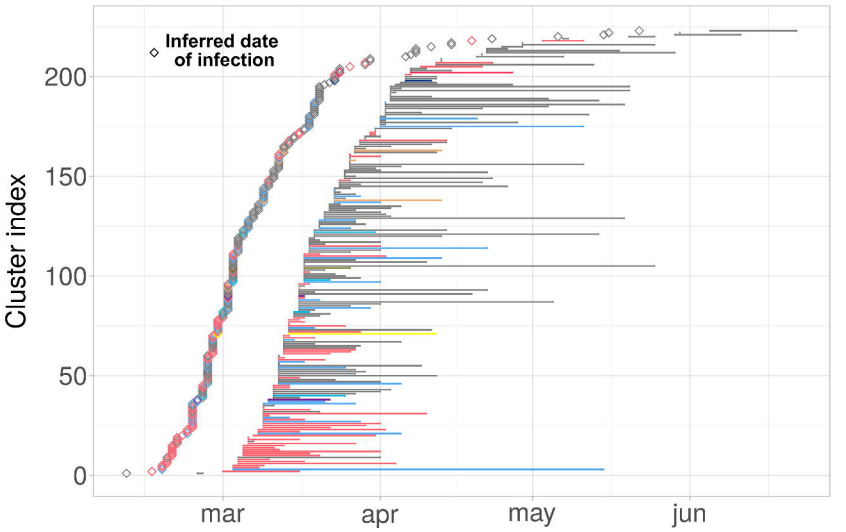

e
Daily introductions
Accumulated introductions
Estimated introductions
500
400
300
200
100
0
Feb-24
Mar-02
Mar-09
Mar-16
Mar-23


f
Slovakia
Norway
Israel
Iceland
Greece
Belgium
Sweden
Australia
Vietnam
Taiwan
Switzerland
Denmark
Scotland
France
China
Portugal
Austria
England
Netherlands
Italy
5
10
15
20
% of potential introductions

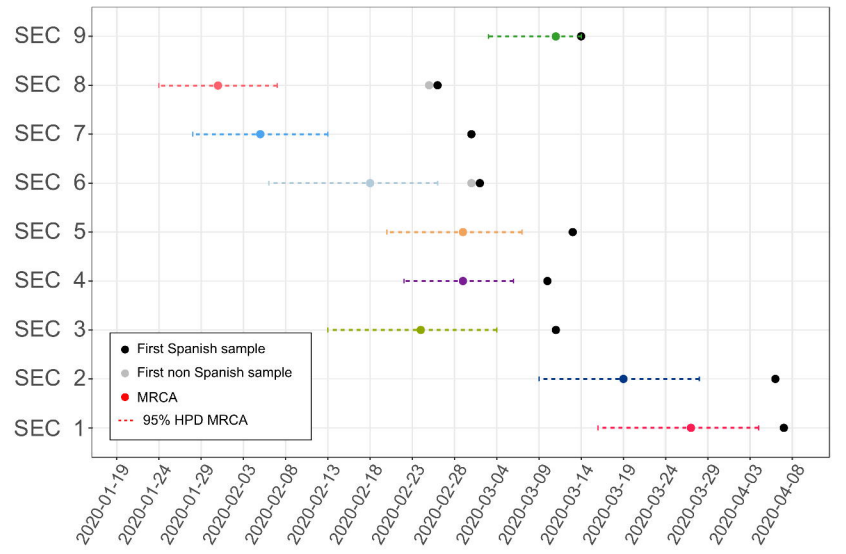
a


0.0001
SEC1
SEC2
SEC3
SEC4
SEC5
SEC6
SEC7
SEC8
SEC9

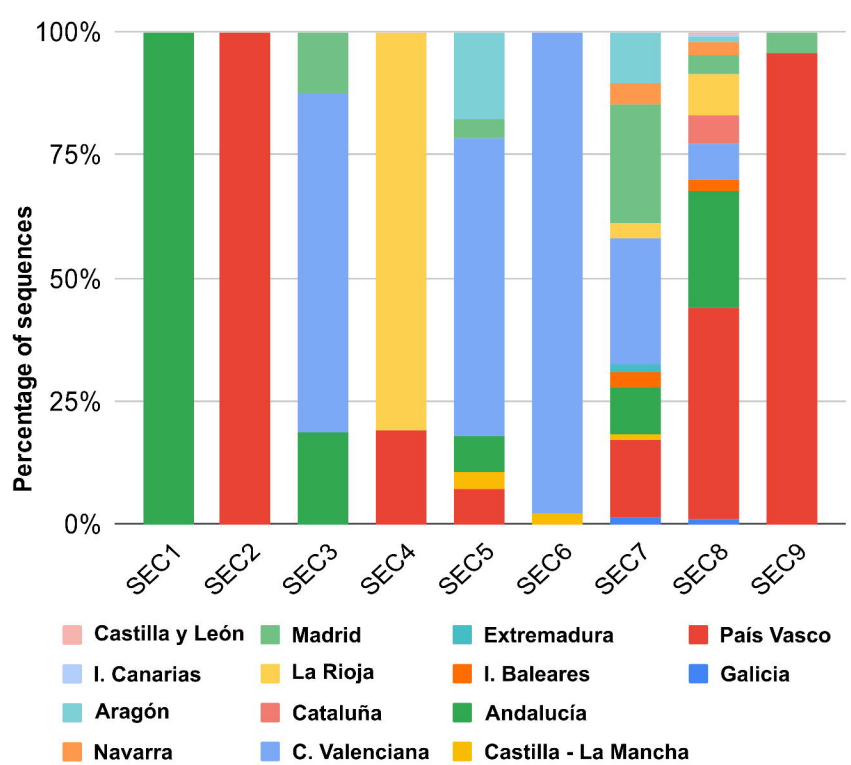
b

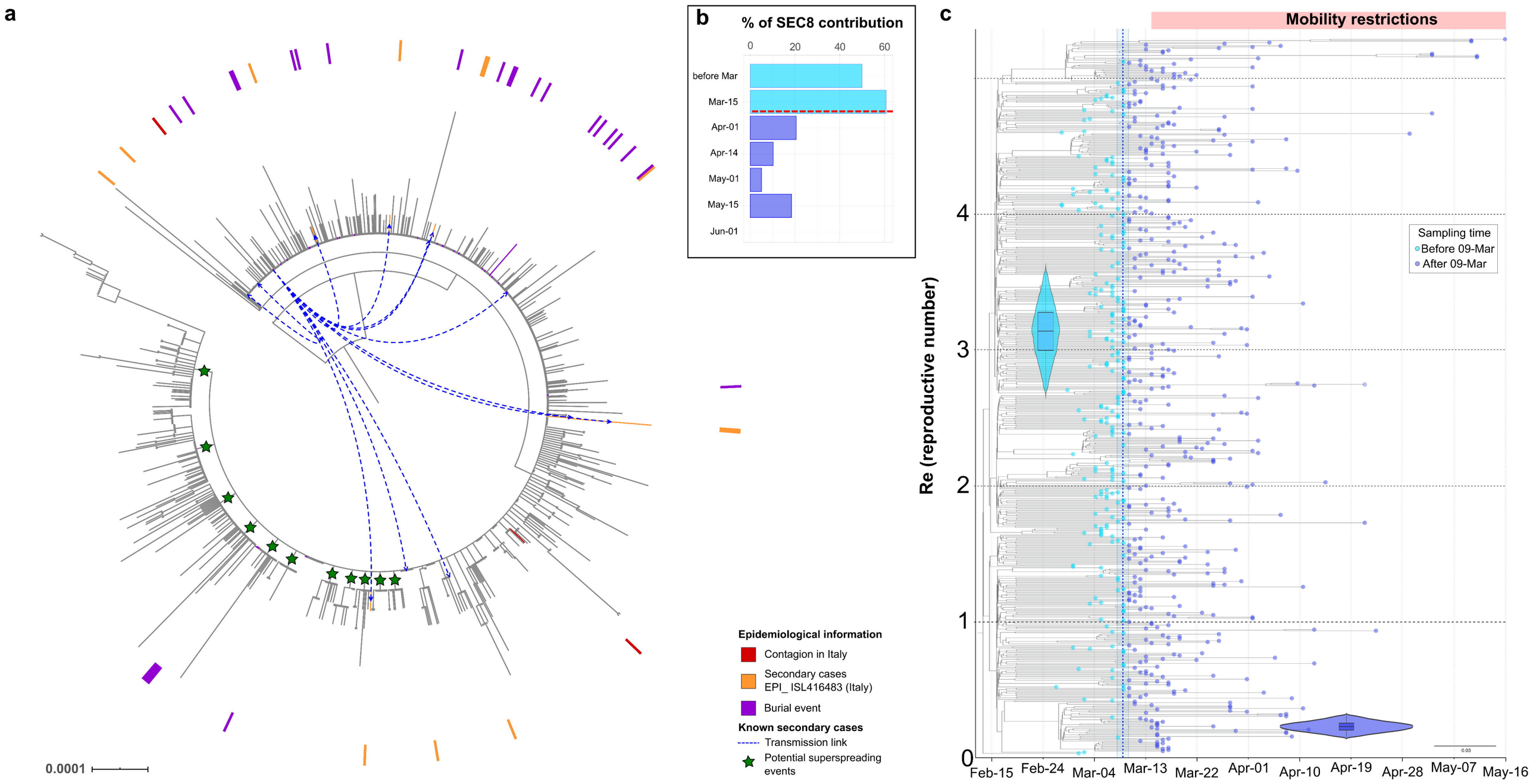

Inferred date of infection
Cluster index
0
50
100
150
200
mar
apr
may
jun

c


SEC 9
SEC 8
SEC 7
SEC 6
SEC 5
SEC 4
SEC 3
SEC 2
SEC 1
First Spanish sample
First non Spanish sample
MRCA
95% HPD MRCA
2020-01-19
2020-01-24
2020-01-29
2020-02-03
2020-02-08
2020-02-13
2020-02-18
2020-02-23
2020-02-28
2020-03-04
2020-03-09
2020-03-14
2020-03-19
2020-03-24
2020-03-29
2020-04-03
2020-04-08

d


Percentage of sequences
0%
25%
50%
75%
100%
SEC1
SEC2
SEC3
SEC4
SEC5
SEC6
SEC7
SEC8
SEC9
Castilla y León
Madrid
Extremadura
País Vasco
I. Canarias
La Rioja
I. Baleares
Galicia
Aragón
Cataluña
Andalucía
Navarra
C. Valenciana
Castilla - La Mancha


a
0.0001
b
% of SEC8 contribution
0
20
40
60
before Mar
Mar-15
Apr-01
Apr-14
May-01
May-15
Jun-01
Epidemiological information
Contagion in Italy
Secondary cases EPI_ ISL416483 (Italy)
Burial event
Known secondary cases
Transmission link
Potential superspreading events
c
Mobility restrictions
Re (reproductive number)
0
1
2
3
4
Sampling time
Before 09-Mar
After 09-Mar
0.03
Feb-15
Feb-24
Mar-04
Mar-13
Mar-22
Apr-01
Apr-10
Apr-19
Apr-28
May-07
May-16