



MINISTERIO
DE CIENCIA
E INNOVACIÓN



INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

INFORME Proyecto COV20/00140 “UNA PERSPECTIVA GENÓMICA DE LA PANDEMIA: LECCIONES EN SALUD PÚBLICA”

Redacción

Iñaki Comas Espadas. Instituto de Biomedicina de Valencia. IBV-CSIC. icomas@ibv.csic.es

Contribuciones

Álvaro Chiner-Oms. Instituto de Biomedicina de Valencia. IBV-CSIC. achiner@ibv.csic.es

Mariana Gabriela López. Instituto de Biomedicina de Valencia. IBV-CSIC. mglopez@ibv.csic.es

Fernando González-Candela. Instituto de Biología Integrativa de Sistemas I2SySBio - Universidad de Valencia – CSIC. fernando.gonzalez@uv.es

En nombre de SeqCOVID - consorcio para la epidemiología genómica de SARS-CoV-2 en España (lista de autores: <http://seqcovid.csic.es/collaborating-institutions/>)



RESUMEN EJECUTIVO

- El consorcio de secuenciación SeqCOVID ha permitido, unido a secuencias depositadas por otras instituciones, colocar a España como el segundo país europeo y cuarto del mundo en número de secuencias generadas del virus.
- La primera ola en España ha sido analizada a la luz de la información genómica de 2170 pacientes de todo el país.
- Los análisis identifican más de 500 introducciones en el país dentro del total de 2170 muestras analizadas. Por tanto, no existe un paciente 0. La mayor parte a partir de mediados de Febrero.
- La primera ola estuvo dominada por hasta nueve genotipos exitosos. En concreto, identificamos un genotipo que representa como media el 30% de nuestras secuencias y llega a representar el 60% las primeras semanas de Marzo. Se caracteriza por una serie de eventos consecutivos que hemos podido reconstruir relativamente bien:
 - **Segunda quincena Febrero.** Entradas múltiples por Valencia (partido de fútbol Atalanta-Valencia) y Madrid (ferias de arte y visita de casos valencianos). Probablemente por más sitios, pero no tenemos acceso a la información epidemiológica detallada.
 - **Finales de Febrero.** Eventos de superdispersión como el de Vitoria. El de Vitoria se corresponde con el brote por un funeral. Llega a población general en País Vasco y La Rioja incluyendo residencia de ancianos. Esto aumenta la frecuencia del genotipo y evita su extinción.
 - **Principios de Marzo.** Movilidad a toda España incluyendo las islas en menos de diez días desde las entradas conocidas. (Figura 1). En la primera semana de Marzo el genotipo representa el 60% de casos de todos los secuenciados.

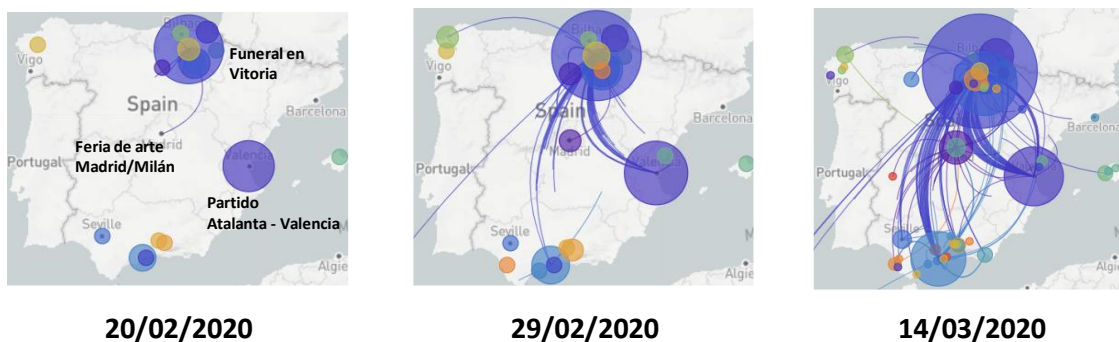


Figura 1. Eventos relacionados con el éxito epidemiológico del SEC8.

- El confinamiento es altamente efectivo con la desaparición de muchos de estos genotipos exitosos.



1. Epidemiología genómica de SARS-CoV-2

En los últimos años se ha dado la confluencia entre dos avances que están cambiando la epidemiología molecular tal y como la conocemos. Por una parte, ahora tenemos acceso al genoma de los microorganismos patógenos a un precio asequible, lo que nos permite usarlos para vigilancia epidemiológica a escala poblacional. El genoma, además, nos da un grado de resolución máximo que, para según qué patógenos, nos permite determinar grupos de transmisión e, incluso, las relaciones entre ellos. Acopladas la posibilidad de secuenciar prácticamente en tiempo real y de unificar datos genómicos y epidemiológicos en los análisis, ha surgido el campo de la epidemiología genómica (Gardy and Loman 2017).

En SARS-CoV-2 estas herramientas se han desplegado como nunca antes y han permitido establecer un mapa de la diseminación del virus por el mundo desde que apareció por primera vez en China hasta hoy. Este mapa está construido por secuencias generadas por grupos de todo el mundo y no es sorprendente que reproduzca las rutas de movilidad de las personas en un mundo global. Además, la resolución del genoma es suficientemente buena como para reconstruir las epidemias a escalas locales. Valgan como ejemplo los estudios llevados a cabo en Reino Unido (Rivett et al. 2020), el llevado a cabo en un hospital de Sudáfrica (“[No Title]” n.d.), para resolver un brote nosocomial que implicó a 119 individuos entre pacientes y personal sanitario, o el desarrollado en el estado de Victoria (Australia) (Seemann et al., 2020) o la ciudad de Nueva York (Gonzalez-Reiche et al. 2020), donde se utilizó para trazar la cadena de transmisión del virus en la región, incluyendo casi un millar de individuos infectados. Recientemente se ha descrito la introducción del coronavirus en Europa y Estados Unidos (Worobey et al. 2020). Además, se ha podido ver cómo eventos de superdispersión locales pueden estar detrás del éxito epidémico en unos sitios y no en otros (Lemieux et al. 2020). Lo verdaderamente importante de este tipo de investigaciones es que son capaces de dotar de información valiosa a las autoridades de Salud Pública, para que puedan adoptar medidas de control de la transmisión del virus.

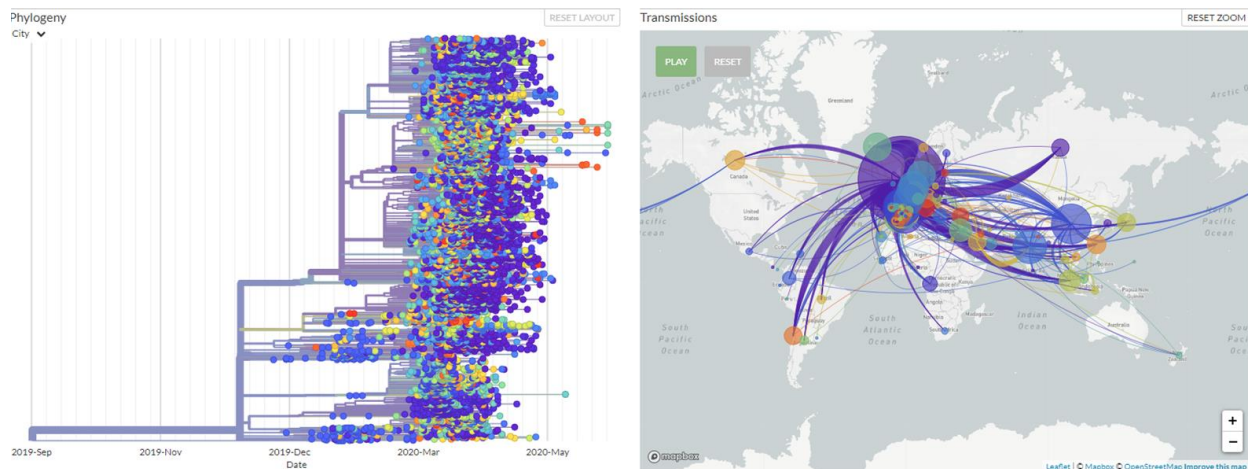


Figura 1. Diversidad filogenética de SARS-CoV-2 y su uso en vigilancia epidemiológica en NextStrain - <https://nextstrain.org/ncov/global>

2. SeqCOVID - Consorcio para la Epidemiología Genómica de SARS-CoV-2 en España

En España se ha organizado un consorcio para usar la secuencia del genoma del coronavirus de veinte mil pacientes para entender cómo el virus llegó a nuestro país y sus patrones de transmisión (seqcovid.csic.es). Actualmente el consorcio está compuesto por 40 grupos de todo el país, la mayoría de ellos radicados en los servicios de Microbiología Clínica hospitalarios. Si bien todavía se está en fase de análisis, queda claro que la epidemia en España no fue iniciada por un “paciente 0”, como era de esperar, y que ha habido un gran intercambio de casos con países de nuestro entorno. Las actualizaciones pueden ser consultadas en: <http://seqcovid.csic.es/>

Actualmente el consorcio cuenta con dos centros principales de secuenciación. El Instituto de Biomedicina de Valencia, perteneciente al Consejo Superior de Investigaciones Científicas, y la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO) perteneciente a la Conselleria de Sanitat de la Generalitat Valenciana. Asimismo, a través del apoyo técnico del consorcio, varios hospitales han adquirido o ampliado la capacidad para secuenciar el SARS-CoV-2. La lista completa de instituciones y personas asociadas se puede consultar en <http://seqcovid.csic.es/collaborating-institutions/> y en el Anexo I de este documento.



MINISTERIO
DE CIENCIA
E INNOVACIÓN



INSTITUT DE
BIOMEDICINA DE
VALÈNCIA



Figura 2. Localización geográfica de instituciones participantes en SeqCOVID. Se puede consultar en <http://seqcovid.csic.es/collaborating-institutions/>.

A fecha 13 de Octubre de 2020 el consorcio ha recibido un total de 7426 muestras de las cuales 4019 han sido secuenciadas y 3800 han sido depositadas en el repositorio público de GISAID (<https://www.gisaid.org/>). En GISAID se centralizan las secuencias producidas por consorcios de todo el mundo. En el caso de España somos el cuarto país del mundo y segundo de Europa con más secuencias contribuidas gracias en gran medida a los esfuerzos de SeqCOVID pero también a otras instituciones que han contribuido al esfuerzo de secuenciación (Figura 3).

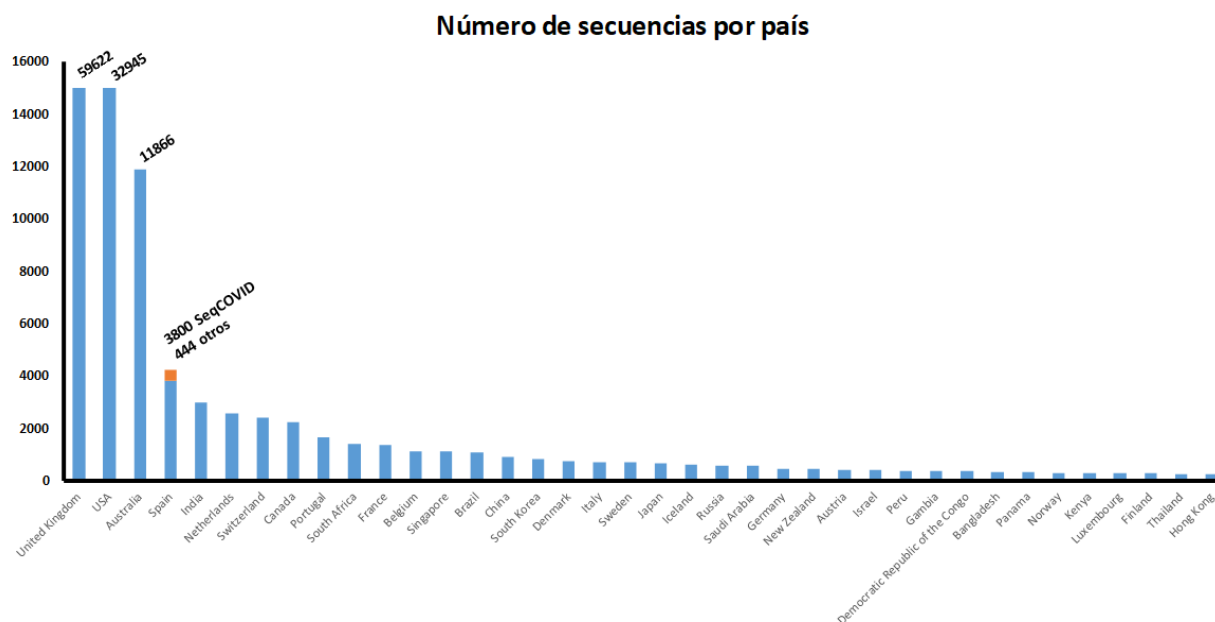


Figura 3. Número de secuencias depositadas en GISAID por país de origen. Para el caso de España incluimos 444 secuencias generadas por otros consorcios y/o instituciones.

3. Diversidad, introducción y dispersión de SARS-CoV-2 en España

El consorcio SeqCOVID ha centrado sus esfuerzos iniciales en entender las primeras etapas del coronavirus en España. Para ello se ha pedido a cada uno de los centros participantes que contribuyeron con muestras representativas de la primera ola de la epidemia y de las primeras semanas del confinamiento. Los análisis que vamos a presentar son el esfuerzo de todos esos grupos, muchos de ellos Unidades de Microbiología Clínica de los hospitales, gracias a los cuales hemos podido obtener una muestra representativa a lo largo de la geografía nacional. Continuamos trabajando para incorporar más centros y más representatividad geográfica en el consorcio.

¿Cuál es la diversidad del virus en España y comparada con otros países?

España presenta una diversidad única del virus cuando se compara con el resto de países de la Unión Europea. Como veremos más adelante, esto se debe a la introducción temprana de linajes que se originaron en China al principio de la epidemia y cuya representación en el resto de países europeos es mucho menor. En España, estos linajes tempranos, han sido los mayores motores de la epidemia inicial.



MINISTERIO
DE CIENCIA
E INNOVACIÓN



ibv INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

La distribución temporal y geográfica de las muestras secuenciadas se puede consultar en (<http://seqcovid.csic.es/microreact/>). Pasamos a describir lo que se observa hasta Septiembre de 2020 siempre teniendo en cuenta que el muestreo de lo que se llama “segunda ola” es todavía limitado.

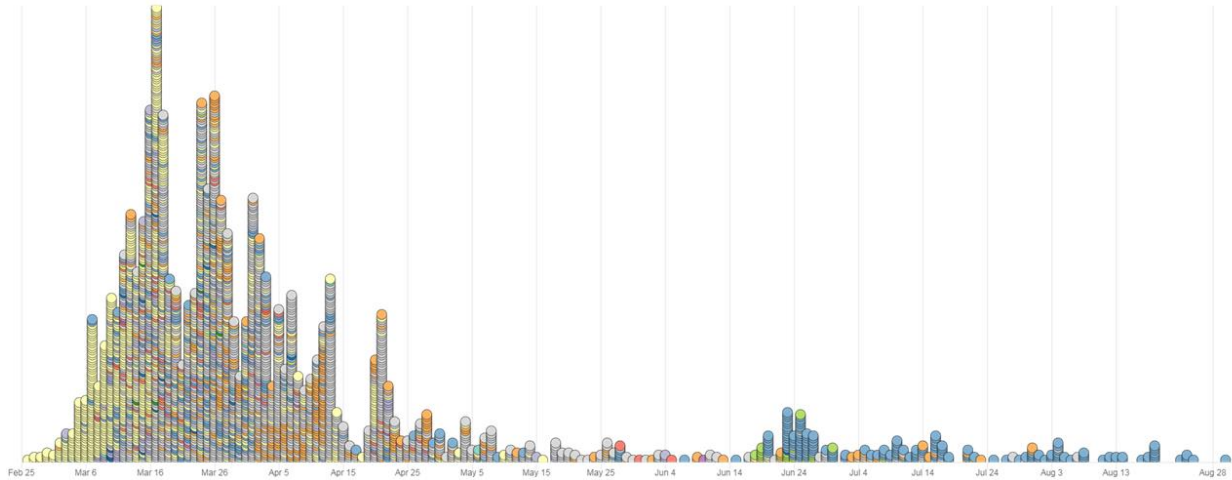


Figure 4. Distribución temporal de las muestras analizadas para este informe. Los colores se refieren a los diferentes linajes circulantes (ver descripción más abajo).

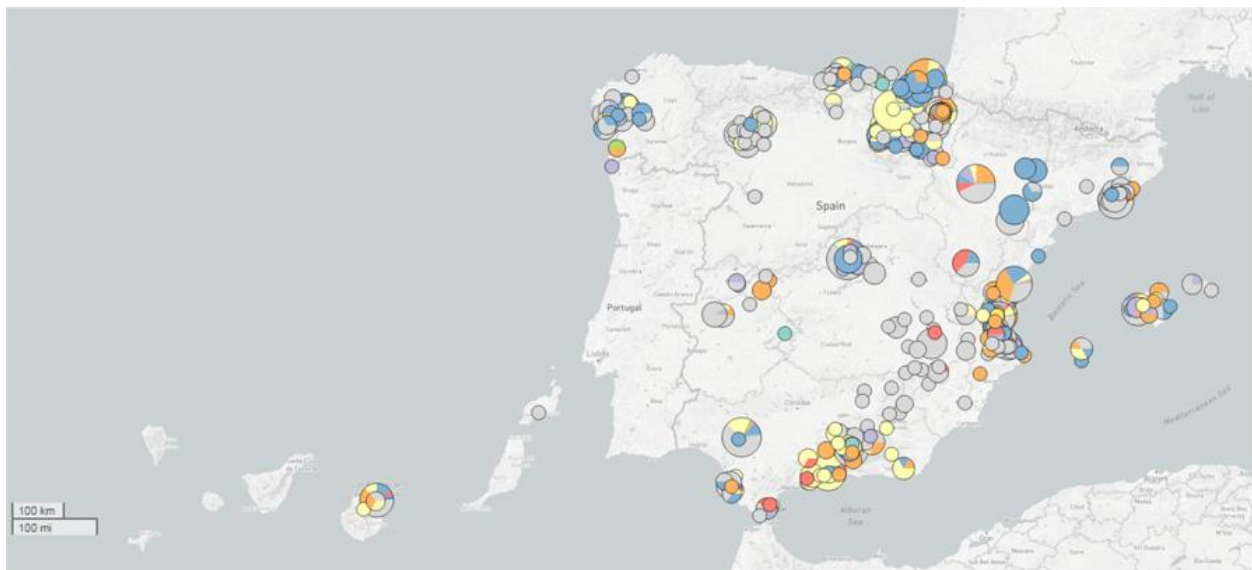


Figura 5. Distribución geográfica de las muestras analizadas en el consorcio. Los colores corresponden a diferentes linajes (ver descripción más abajo).

Los diferentes linajes siguen la nomenclatura propuesta por el consorcio de secuenciación inglés y descrito en (Rambaut A et al. 2020). Aunque existen otros esquemas que se comparan en la siguiente publicación en la que ha participado el consorcio (Alm et al. 2020). Si bien los linajes no nos dejan ver el grano fino de la dispersión del virus en España sí nos dan una idea de la diversidad y de la importancia temporal de cada uno de ellos. Por ejemplo, el inicio de la epidemia se vio dominado en España por linajes A. Estos linajes son descendientes directos de los primeros que aparecieron en China y representan una diferencia importante con el resto de epidemia de otros países, dominados por linajes más tardíos. Esta información será importante más adelante en el análisis detallado de la epidemia en España.

Figure S3. Distribution of the most common lineages (cov-lineages.org) in the WHO European Region, based on all high-quality genomes in GISAID until 10 July 2020.

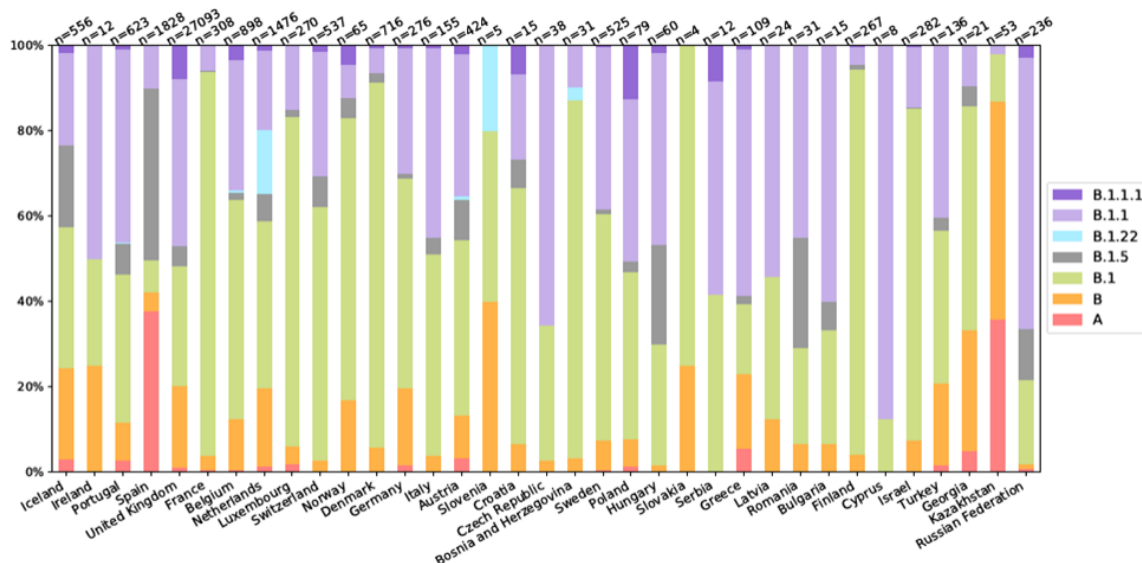


Figura 6. Distribución de linajes por país. Se aprecia que en España casi el 40% de las secuencias pertenecen a linajes A. (Alm et al. 2020 EuroSurveillance)

La dinámica temporal en España de cada linaje refleja también la importancia relativa de los linajes A comparada con la del Reino Unido, por ejemplo.

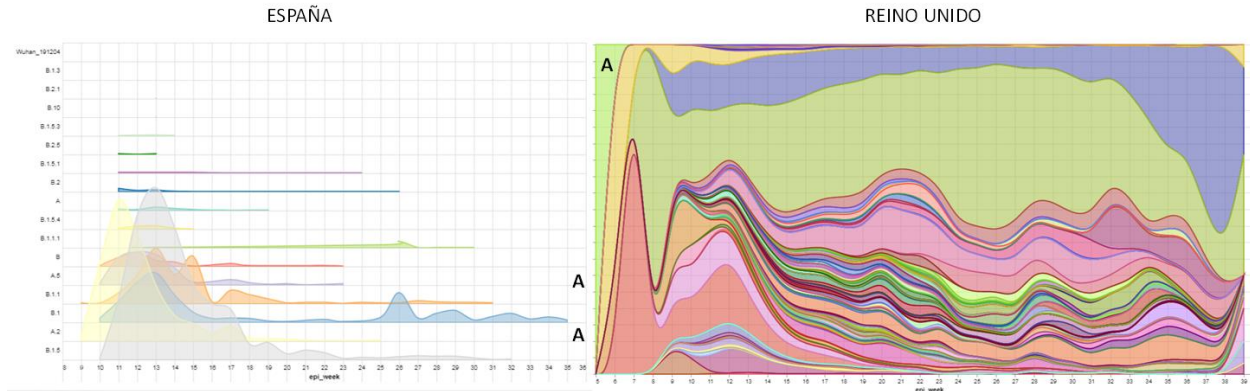


Figura 7. Los linajes A tuvieron una mayor representación en la primera ola en España comparada con el Reino Unido. Una situación parecida ocurre con el resto de países de Europa.

Múltiples introducciones contribuyeron al éxito de la epidemia en España

Para los análisis de la primera ola nos basaremos en 2170 secuencias muestreadas principalmente entre los meses de Febrero y Mayo. En la figura se ven los casos confirmados por el Ministerio de Sanidad y los casos secuenciados y que se han podido incluir en el análisis.

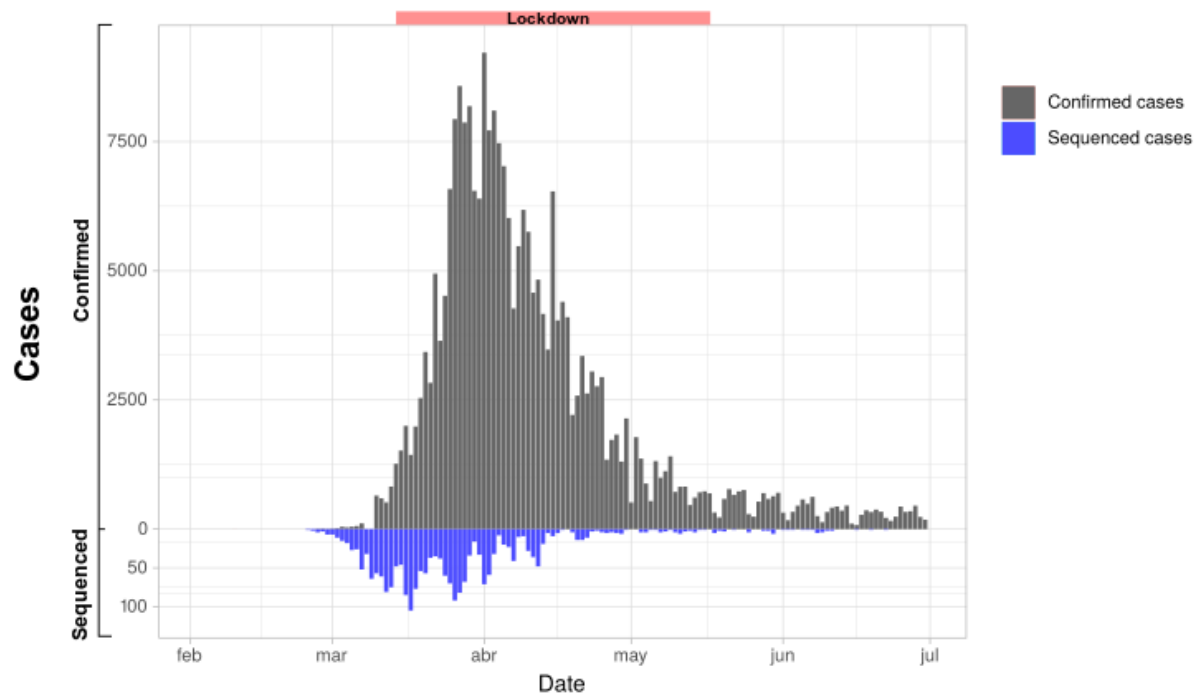


Figura 8. Panel superior, número de casos confirmados de COVID-19. Panel de casos inferior, número de casos secuenciados.

El caso más antiguo de COVID-19 en España data de finales de enero. Se trata de un paciente que realizó un viaje a Nepal y, tras su regreso a España el 30 de enero, falleció (13/02/2020) a causa de una neumonía de origen desconocido, aunque con sospecha de COVID-19. El caso fue identificado en una autopsia post-mortem ante la sospecha de que el cuadro que padecía el paciente pudiera ser causado por SARS-CoV-2. La carencia de vínculo directo con China impidió la realización de pruebas diagnósticas para la infección por el virus, al no cumplir con los criterios establecidos por el Ministerio de Sanidad en ese momento. Sin embargo, la identificación de un caso temprano no implica que el paciente en cuestión sea el “generador” o “iniciador” de la pandemia en España. De hecho, ni en los estudios epidemiológicos ni en nuestros análisis este paciente no generó ningún caso secundario. Como se ha observado en otros países, los resultados de nuestros análisis filogenómicos sugieren la existencia de múltiples introducciones independientes del virus en el país.



Tree scale: 0.0001

Spanish_sequences
■ Spanish_sequences

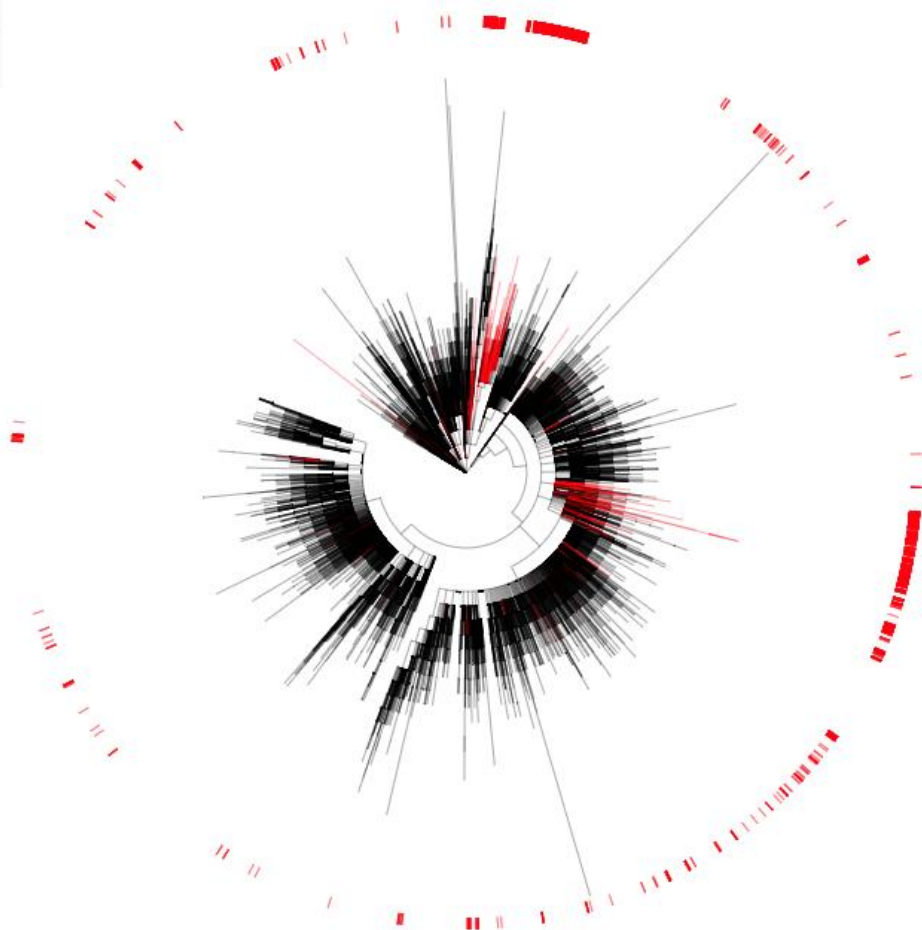


Figura 5. Múltiples introducciones del coronavirus en España cuando se pone en un contexto global.

En este sentido, puede observarse en la figura 5 que existen numerosos casos aislados (512) al igual que grupos compuestos solo por muestras españolas (224) o grupos de muestras españolas altamente relacionadas pero que no forman grupos únicos por estar incluidas muestras de otros países entre ellas (30), a lo largo de toda la filogenia conocida del virus. Los 224 grupos con solo muestras españolas concentran aproximadamente el 40% (827 casos) del total de las muestras analizadas hasta la fecha (2170 casos). Muchas de estas agrupaciones corresponden, probablemente, a brotes de transmisión local, siendo imperativo validarlos con información epidemiológica. Aun considerando que no se trata de un estudio poblacional y que el tamaño de la muestra es reducido, el número de casos incluidos en estos brotes potenciales

es elevado. La distribución está en concordancia con nuestro conocimiento sobre la transmisión del coronavirus (Figura 6). Si bien hay muchos grupos de dos o tres casos, también es muy común encontrar grupos de más de cinco, incluso de hasta 23 (Figura 6). Este dato indica que los eventos de super dispersión tienen un gran impacto en el éxito de la pandemia, como veremos más adelante.

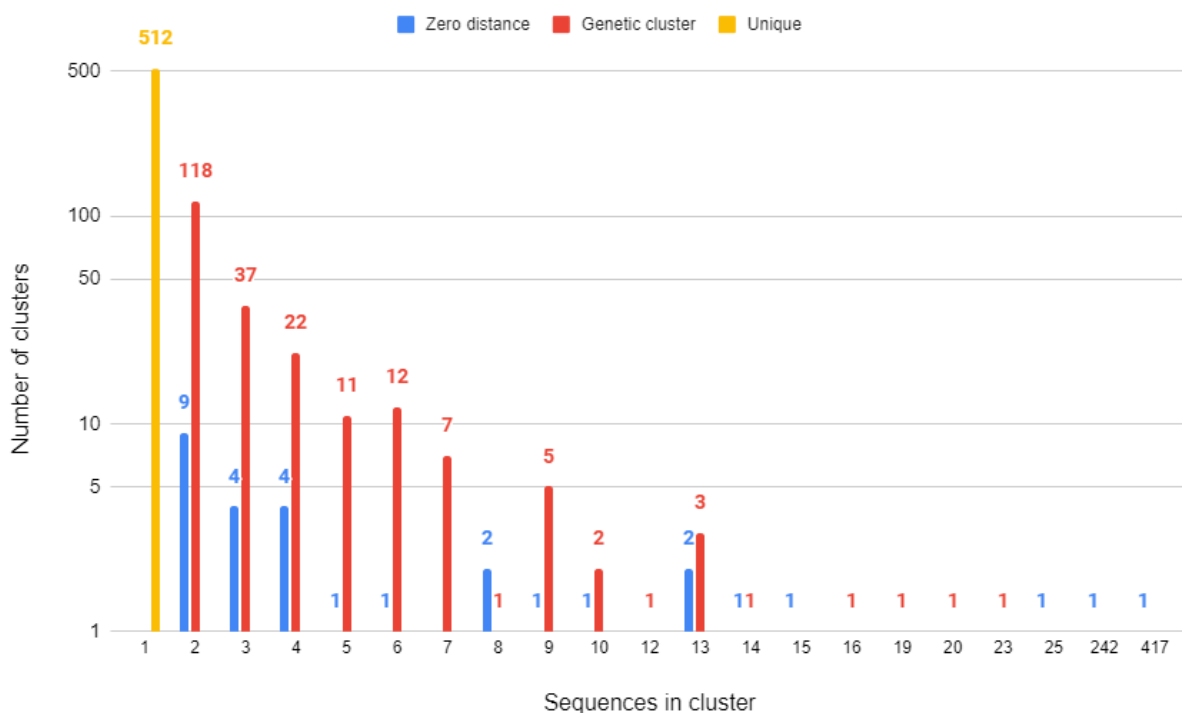


Figura 6. Número de agrupaciones detectadas en el dataset. Consideramos que dos muestras están agrupadas si comparten clado filogenético y no hay muestras de otros países. Son, por tanto, muestras estrechamente relacionadas entre sí y del resto de la filogenia. Por lo tanto, pueden representar grupos de transmisión locales si bien se tendría que confirmar con la información epidemiológica.

La presencia de casos únicos del principio de la epidemia así como de agrupaciones claras, son indicativos del número mínimo de introducciones en el país durante Febrero y Marzo. Para determinarlo analizamos cuántos grupos habían ocurrido antes del 14 de Marzo, fecha del estado de alarma y en la que las fronteras españolas se cerraron.

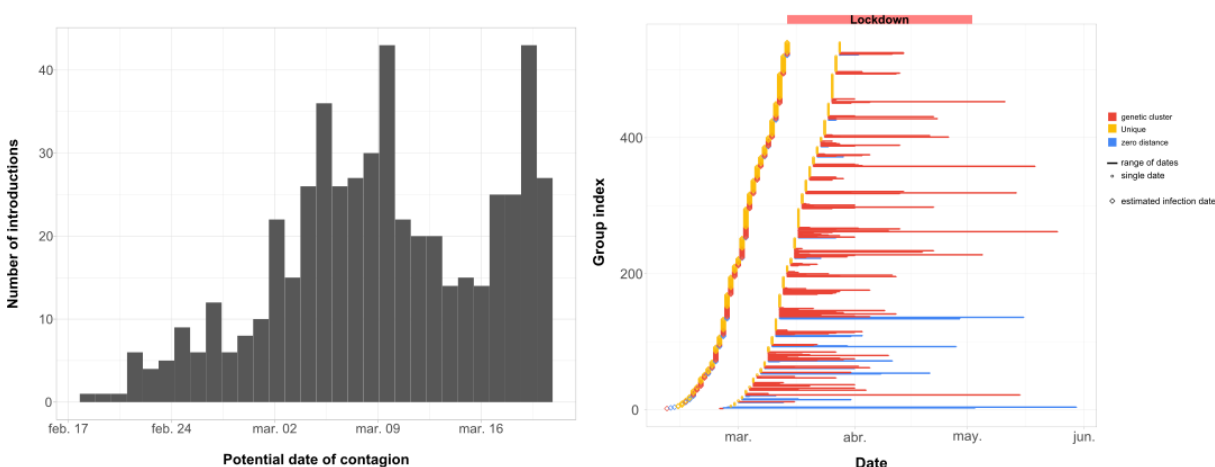


Figura 7. Izquierda. Fecha de contagio en el extranjero probable para cada caso que introdujo el virus en España. Se basa en restar 14 días a la fecha diagnóstico a cada caso represente de la introducción. Por lo tanto, la introducción podría haberse podido haberse dado en un rango desde el momento marcado en el histograma hasta catorce días después. **Derecha.** Primer caso asociado a cada grupo genético y su fecha probable de infección basada en un tiempo de incubación máximo de 14 días. Se muestran solo los grupos cuya fecha probable de infección se dió antes del inicio del estado de alarma y por tanto compatibles con introducciones del exterior.

Basado en nuestro análisis, un total de 191 grupos genéticos se iniciaron antes del inicio del confinamiento. A estos hay que añadir 328 casos únicos que no están dentro de ningún grupo español y que también se originaron antes del confinamiento. En total podemos definir un total de 519 introducciones. Este número debe ser tomado con cautela, pues el número de secuencias analizadas es mucho menor que el número de casos y, por tanto, representa un valor mínimo. Sin embargo, indica el gran número de importaciones que, con mayor o menor éxito, contribuyeron a la epidemia. Atendiendo a las fechas en las que se dieron las exposiciones las introducciones se aceleraron a partir del 24 de Febrero (Figura 7) y se mantuvieron constantes casi hasta el estado de alarma.

4. Impacto de las grandes ciudades en la dispersión del virus sobre su área metropolitana y de influencia/Las grandes ciudades como centros de dispersión del virus

El impacto que han tenido las grandes ciudades como centro de dispersión/transmisión del virus hacia las zonas metropolitanas y sus áreas de influencia ha sido evaluado en la Comunidad Valenciana, en el marco de un proyecto local que posibilitó un esfuerzo de muestreo mayor. El razonamiento aplicado es que si las grandes ciudades han sido centros de dispersión del virus hacia las zonas periféricas, entonces esperamos una máxima diversidad genética del virus en la ciudad de Valencia y área metropolitana que se va reduciendo con la distancia geográfica. Este

resultado sería compatible con un modelo de aislamiento por distancia, donde sólo parte de la diversidad de la población original (Valencia) llega a sus alrededores y es compatible con los patrones de movilidad poblacional registrados en las grandes ciudades antes del confinamiento.

Nuestro análisis revela una fuerte influencia del área metropolitana de Valencia en el área de influencia periférica (Figura X), que llega hasta los 70 Km del centro de la ciudad. En este punto, otras áreas de influencia, como Gandía o Castellón, impactan sobre la diversidad del virus.

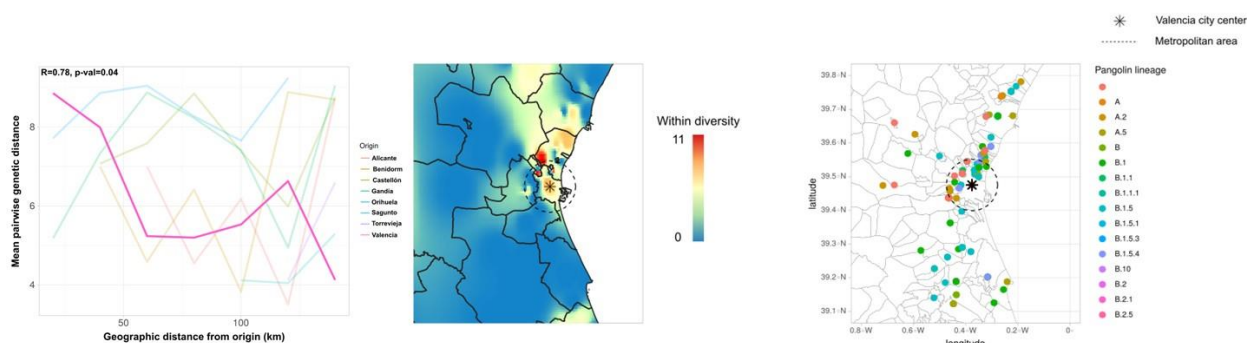


Figure 8. A) Reducción de la diversidad genética entre muestras del virus conforme nos alejamos del centro de Valencia. El eje de la X marca los Km de distancia. La diversidad cae hasta los 50 Km y se estabiliza hasta los 70 Km. Por otra parte el mismo análisis no es significativo para otras grandes urbes cuyo impacto en su área de influencia es mucho menos. B) Mapa de calor de la diversidad en Valencia (asterisco), área de metropolitana (círculo) y alrededores. Colores amarillos y rojos indican más diversidad mientras que colores azules indicando muy baja diversidad en esas zonas. D) distribución de linajes según localidad analizada.

5. Clados epidémicos causantes de la mayoría de los casos en España

De la Figura 5 queda claro no sólo que hubo múltiples introducciones sino que, además, muchas muestras se concentran en diferentes zonas de la filogenia. Esto sugiere que unos pocos genotipos del virus generaron un gran número de casos secundarios. A diferencia del paciente “0”, estos clados serían los responsables de la pandemia en España (Figura 9).



Tree scale: 0.0001

SECS

- SEC1
- SEC2
- SEC3
- SEC4
- SEC5
- SEC6
- SEC7
- SEC8
- SEC9

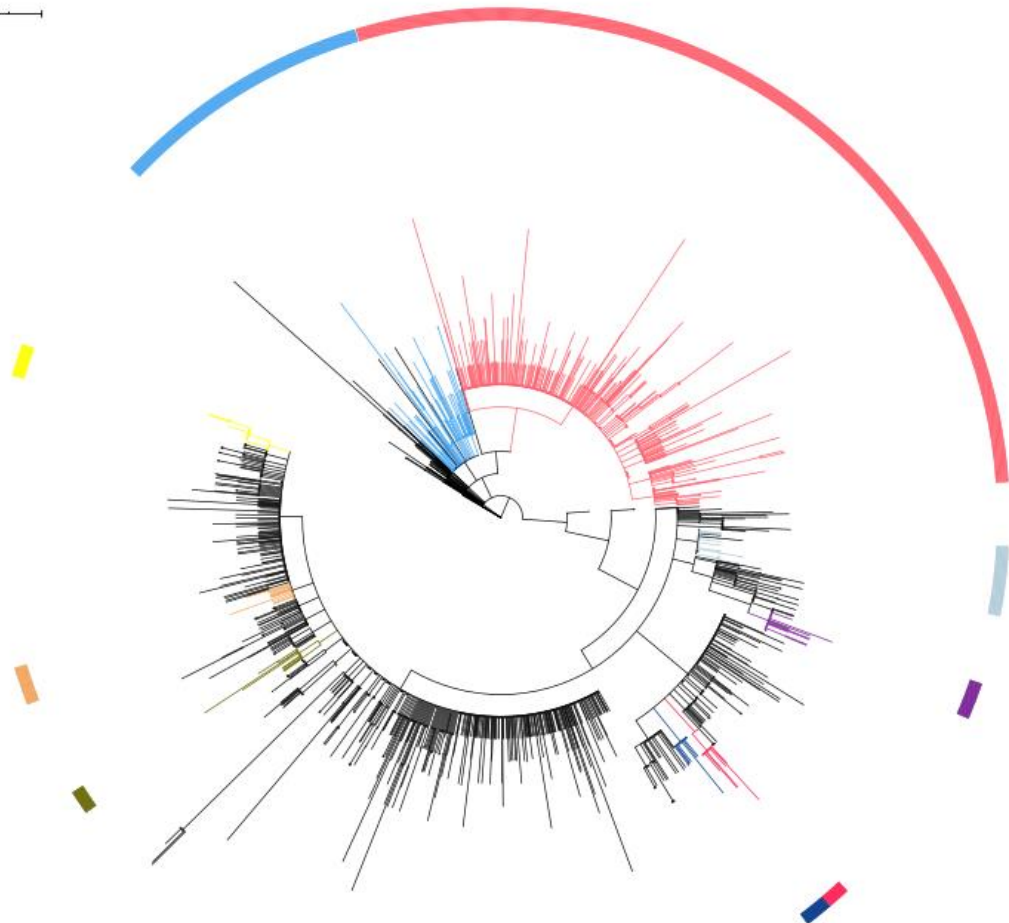


Figura 9. Filogenia de las muestras de origen español coloreada según el SEC al que pertenecen.

Para probar esta hipótesis identificamos en una filogenia global con 32914 muestras de todo el mundo, clados con buen soporte estadístico y con una mayoría de muestras españolas.

Identificamos nueve clados que pasamos a denominar “Spanish epidemic clases” o SEC.

Total samples in SEC and Spanish samples in SEC

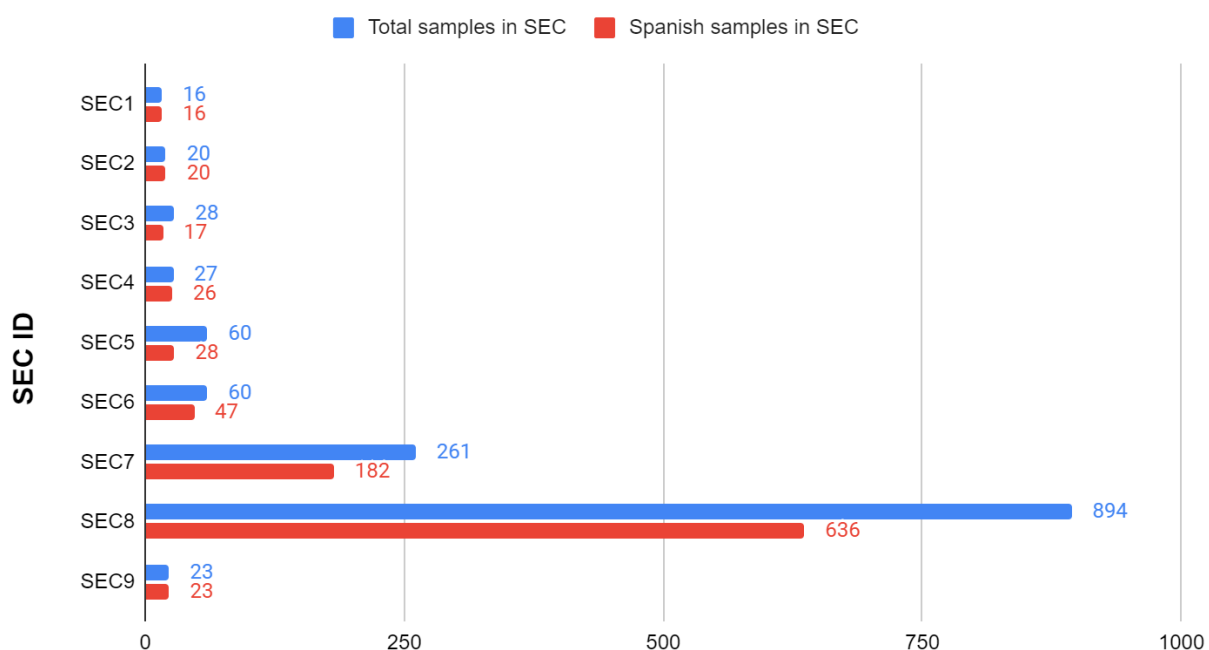


Figura 10. Número total de muestras en cada SEC y número total de muestras españolas.

Lo primero que llama la atención es que dos SEC, SEC8 y SEC7, incluyen el 30% y al 10%, respectivamente, de todas las muestras españolas. Esto nos indica que la introducción de dos genotipos concretos explica una parte importante de toda la epidemia en el país. Los SEC8 y SEC7, además, son parte de los linajes A. Esto sugiere que las introducciones se dieron muy pronto en España pues esos linajes fueron los primeros en aparecer en Asia mientras que en Europa son los linajes B los que predominan. En el siguiente apartado veremos el posible origen de alguno de los SEC.

El éxito de cada SEC puede ser debido a diferentes factores. Por ejemplo, se pueden haber expandido localmente, provocando un gran número de casos, pero no dispersados por el resto de la geografía española. Para estudiar estos patrones analizamos la distancia geográfica media entre muestras de un mismo SEC. La mayoría de los SEC, sobretudo los más pequeños, apenas se han dispersado, están la mayoría de las muestras en un rango de unos 0-58 km entre sí (SEC1-SEC6, SEC9 - Figura 11). Por el contrario, la distancia media en SEC7 es 323 km (rango intercuartil 225-447) y en SEC8 de 371 km (rango intercuartil 76-634). Esto indica que particularmente el SEC8 tuvo una rápida dispersión por todo el país después de introducirse. De hecho, como se puede ver en la Figura X, el rango geográfico es mucho mayor que en el resto

de SEC sin que lleve correlacionado una mayor acumulación de mutaciones. Ello nos indica que los SEC7 y SEC8 se expandieron rápidamente por gran parte de la península, lo que se corrobora en los siguientes análisis.

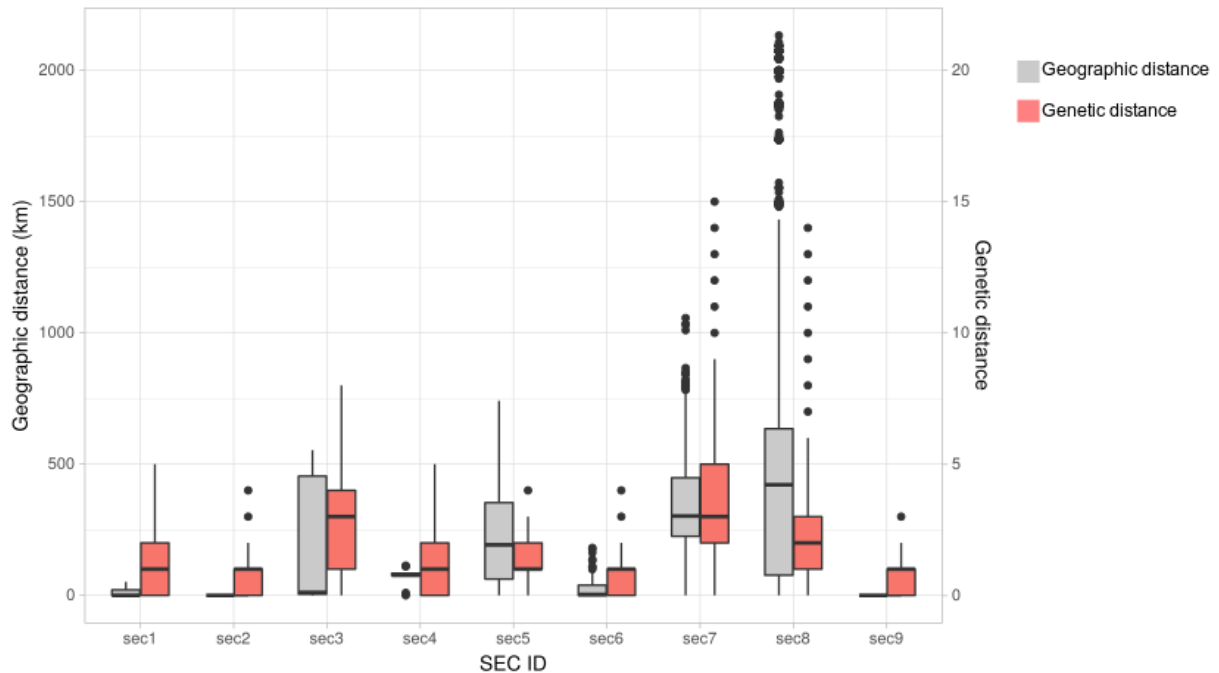


Figura 11. Distancia genética y distancia geográfica entre muestras de un mismo SEC.

6. Origen temporal de los Clados Endémicos Españoles (SECs)

Los SEC7 y SEC8 son los principales grupos causantes de la epidemia en España, ya que concentran el mayor número de casos secundarios, es decir de contagios. Ambos corresponden a grupos filogenéticamente basales asociados a las cepas circulantes en China.

El resultado del análisis de datación bayesiano (Tabla 1, Figura 11), que aprovecha la acumulación de cambios genéticos del virus a lo largo del tiempo, nos indica que en nuestra muestra, las introducciones más tempranas y exitosas del virus, correspondientes al SEC7 y al SEC8 ocurrieron durante el mes de Febrero. El período de introducción señalado coincide temporalmente con el primer paciente diagnosticado en España y corrobora que durante la segunda quincena de Febrero el virus se expandió rápidamente. No puede descartarse que el virus hubiera podido ingresar antes en el país, dado que el muestreo al inicio de la pandemia fue reducido y nuestra datación se corresponde con la evidencia epidemiológica disponible. Sin embargo, si hubo introducciones previas, no ha quedado rastro posterior de las mismas en la epidemia. El origen de los otros SECs se sitúa entre finales del mes de febrero y el mes de marzo y en la mayor parte de los casos representan brotes grandes pero locales.

Tabla 11. Origen temporal de los SECs obtenido mediante un modelo bayesiano coalescente. Se indica la primera y la última muestra española disponible, el origen estimado y el intervalo de confianza del mismo (CI)

SECs	PRIMERA MUESTRA	ULTIMA MUESTRA	ORIGEN DEL ANCESTRO	ORIGEN 95% CI	Rango estimado de introducción
SEC1	2020-04-07	2020-04-27	2020-03-27	[2020-03-16;2020-04-04]	[2020-03-27;2020-04-07]
SEC2	2020-04-06	2020-04-11	2020-03-19	[2020-03-09;2020-03-28]	[2020-03-09;2020-04-068]
SEC3	2020-03-11	2020-04-01	2020-02-24	[2020-02-13;2020-03-04]	[2020-02-24;2020-03-11]
SEC4	2020-03-10	2020-04-03	2020-02-29	[2020-02-22;2020-03-06]	[2020-04-03;2020-03-10]
SEC5	2020-03-13	2020-04-10	2020-02-29	[2020-02-20;2020-03-07]	[2020-04-10;2020-03-13]
SEC6	2020-03-02	2020-03-31	2020-02-18	[2020-02-06;2020-02-26]	[2020-03-31;2020-03-02]
SEC7	2020-03-01	2020-05-15	2020-02-05	[2020-01-28;2020-02-13]	[2020-05-15;2020-03-01]
SEC8	2020-02-26	2020-05-16	2020-01-31	[2020-01-24;2020-02-07]	[2020-05-16;2020-02-26]
SEC9	2020-03-014	2020-04-12	2020-03-11	[2020-03-03;2020-03-14]	[2020-04-12;2020-03-014]

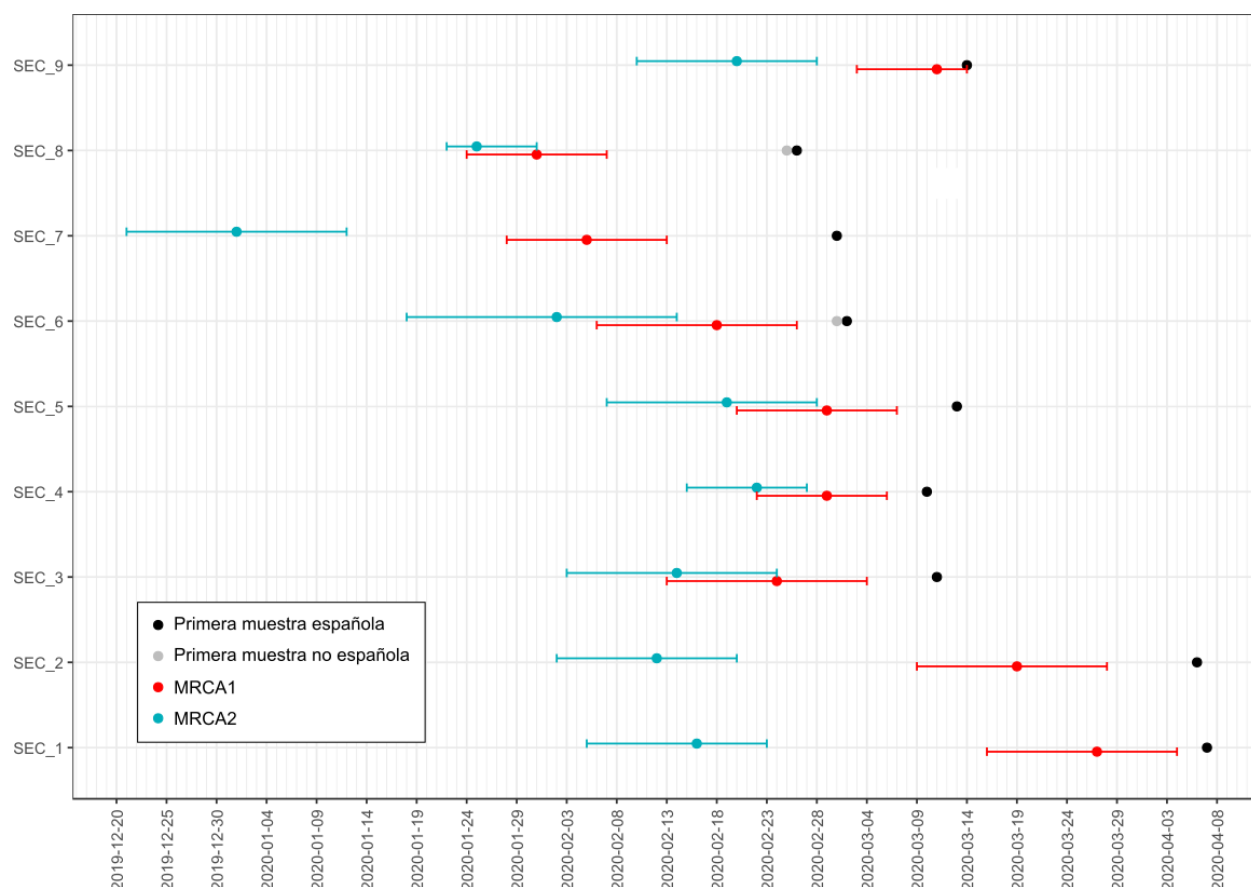


Figura 12 Origen temporal de los SECs estimado mediante un análisis bayesiano coalescente. MRCA1 indica el ancestro más reciente del grupo que probablemente no estaba todavía en España mientras que el punto negro indica la primera muestra secuenciada en España. Esas dos fechas nos dan el rango de tiempo en el que se produjeron las primeras introducciones de cada grupo en España. Como se puede observar, los SEC7 y 8 son los más tempranos mientras que los SEC1 y SEC2 probablemente ocurrieron justo antes del confinamiento.

7. Impacto del confinamiento sobre el control epidémico

Se evaluó el efecto del confinamiento en la tasa de crecimiento de la pandemia considerando aquellos grupos epidémicos españoles que concentran el mayor número de casos y, por tanto, son los mayores grupos de dispersión en el país. Se evaluó el impacto de las medidas de confinamiento en los SECs 7 y 8. Para ello se evaluó si existió un cambio en el R_e y cuándo se produjo el mismo mediante modelos bayesianos de nacimiento-muerte (*birth-death*). La variable R_e representa el cambio a lo largo de la epidemia del número reproductivo efectivo, o número de infecciones secundarias causadas por cada persona infectada, en un área geográfica



determinada. Su valor nos informa directamente de si la epidemia está en fase de crecimiento, decrecimiento o estabilización. Lo primero que destaca es un alto valor de R_e , de alrededor de 2.5 (Figura 13, 14). Este resultado está en consonancia con lo que sabemos de la R_e del SARS-CoV-2 al principio de la epidemia, cuando se aproximaba a su R_0 ([Standl et al. 2020](#)).

El SEC7 registra una disminución significativa en el R_e , prácticamente hasta 0.5, entre el 15 y 24 de marzo, en concordancia con la fecha de confinamiento. En la filogenia de la figura 13 puede observarse que el mayor número de contagios, que se observan como puntos de bifurcación filogenética, se produce antes del inicio del confinamiento. En el caso del SEC8, la fecha de cambio del R_e es ligeramente anterior al confinamiento, entre el 8 y 10 de marzo donde las primeras restricciones parciales en diferentes partes de España ya estaban funcionando. Nuestro análisis, basado únicamente en la diversidad muestreada del virus en cada momento, indica que el confinamiento fue altamente eficiente y ayudó a evitar un gran número de casos y de fallecimientos.

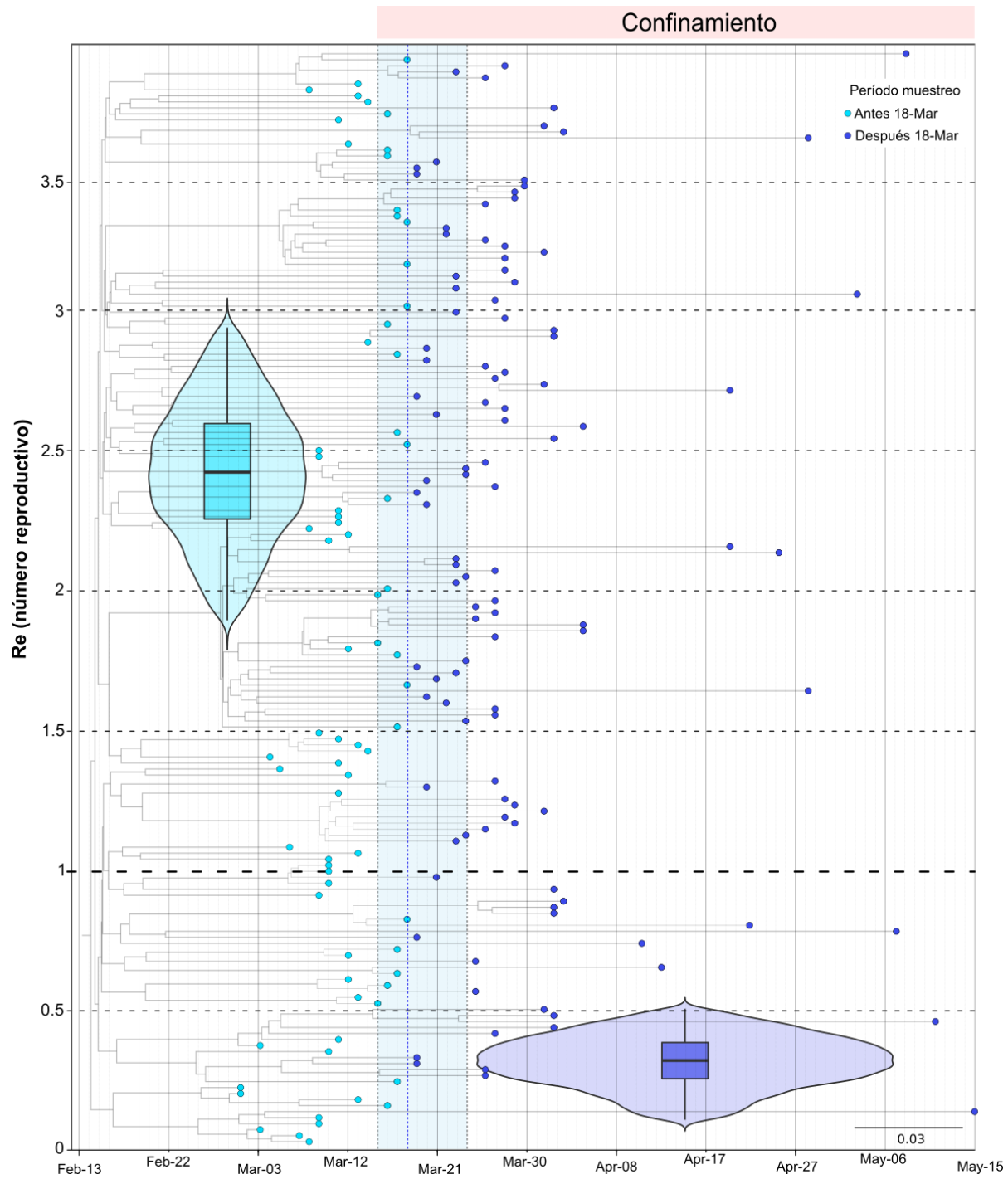
Figura 13 Estima del cambio en número reproductivo (R_e) en el SEC7 mediante modelos bayesianos de nacimiento-muerte (BDSKY). El modelo permite un punto de cambio durante el período de muestreo. El eje X representa las fechas desde el origen estimado del SEC7 a mediados de febrero de 2020 (95% HPD: 07-22 de febrero) hasta la fecha de la última muestra el 15 de mayo. La línea azul representa la fecha del cambio más significativo en el R_e , 18 de marzo (95% HPD: 15–24 marzo). El eje Y representa los valores de R_e , los gráficos de violín representan la distribución posterior para ese parámetro antes y después del cambio, con una mediana de 2.42 (95% HPD: 1.89–2.93) y 0.3 (95% HPD: 0.11–0.41), respectivamente. Los terminales del árbol filogenético estimado han sido coloreados de acuerdo a la fecha de muestreo antes o después del día de cambio (18 de marzo).



MINISTERIO
DE CIENCIA
E INNOVACIÓN



ibv INSTITUT DE
BIOMEDICINA DE
VALÈNCIA





MINISTERIO
DE CIENCIA
E INNOVACIÓN



ibv INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

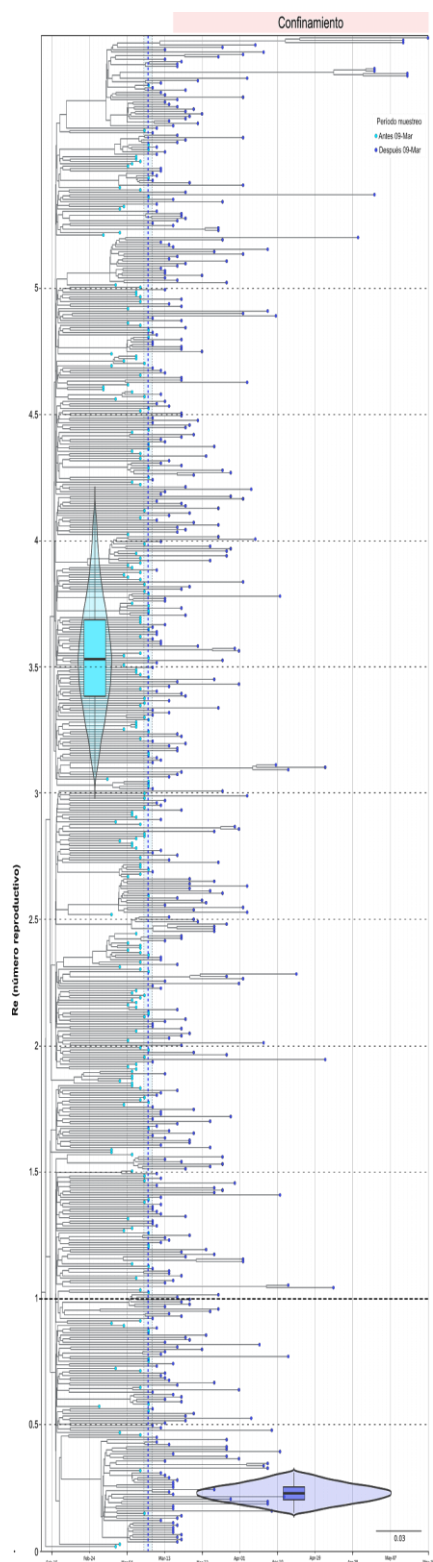


Figura 14 Estima del cambio en número reproductivo (R_e) en el SEC8 mediante modelos bayesianos de nacimiento-muerte (BDSKY), el modelo permite un punto de cambio durante el período de muestreo. El eje X representa las fechas desde el origen estimado del SEC8 a mediados de febrero de 2020 (95% HPD: 05-19 de febrero) hasta la fecha de la última muestra el 16 de mayo. La línea azul representa la fecha del cambio más significativo en el R_e 9 de marzo (95% HPD: 8–10 marzo). El eje Y representa los valores de R_e , los gráficos de violín representan la distribución posterior para ese parámetro antes y después del cambio, con una mediana de 3.56 (95% HPD: 3.06–4.09) y 0.23 (95% HPD: 0.15–0.31), respectivamente. Los terminales del árbol filogenético estimado han sido coloreados de acuerdo a la fecha de muestreo antes o después del día de cambio (9 de marzo).

8. Éxito epidemiológico del SEC8

Saber exactamente a qué se debió el éxito de unas variantes y no de otras es siempre complejo. Sin embargo, al principio de una epidemia rara vez una mutación está asociada a éxito epidemiológico o virulencia y son más probables eventos fortuitos, particularmente eventos fundadores. Es lo que creemos que ha ocurrido con el SEC8. Para empezar un análisis de la contribución de cada SEC a la epidemia en España indica que aquellos que se introdujeron antes fueron los que más casos causaron (Figura 15). En el caso del SEC8 esa contribución llegó a ser del 60% de los casos secuenciados antes del confinamiento (Figura 15).

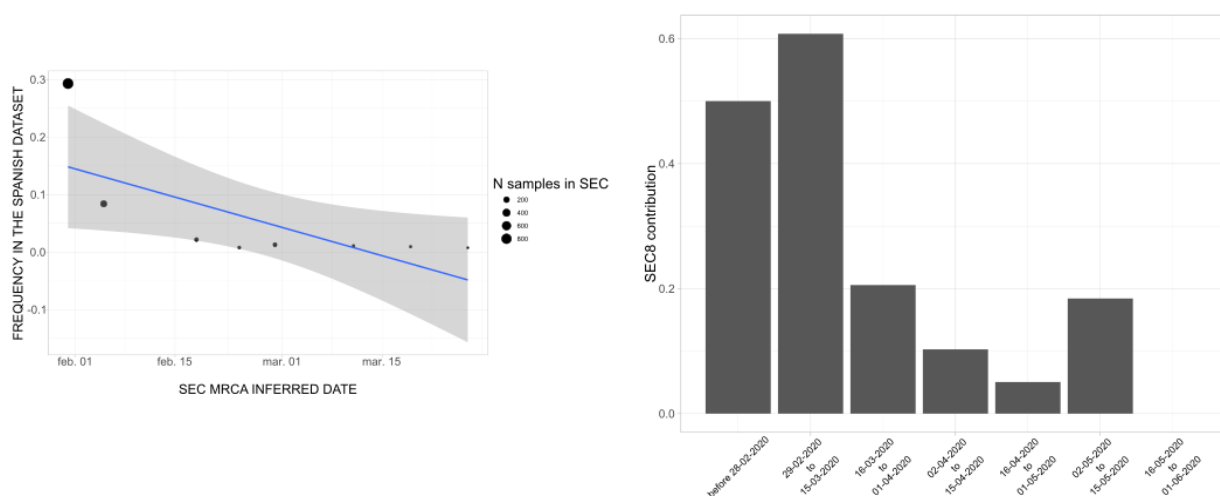


Figura 15. Izquierda. Representa la contribución de cada SEC al total de muestras españolas dependiendo de su fecha más probable de origen. **Derecha.** Representa la contribución del SEC8 al total de muestras secuenciadas en el tiempo.



Este dato refuerza la idea de que mucho del éxito epidémico depende de haberse introducido antes de que hubiera medidas en funcionamiento. Sin embargo, las importaciones individuales pueden no tener ningún efecto en la epidemia. El primer caso conocido en España se importó de Nepal/China y no dejó rastro. Por lo tanto, el éxito posterior debe estar asociado a otras variables. Solo con la secuencia genómica es difícil reconstruir lo que pasó y qué variables fueron relevantes. Sin embargo, cruzado con datos epidemiológicos detallados podemos hacernos una idea de cómo unos pocos casos pueden convertirse en miles en solo unas semanas. Para el SEC8 tenemos datos detallados sobre algunos casos y eventos que parecen haber estado en la base de su éxito.

Por una parte, el SEC8 fue introducido repetidamente en España al menos por dos vías, Madrid y Valencia (marcadas en Figura 16). Existe constancia epidemiológica que algunos de los casos secuenciados están asociados al partido Atalanta-Valencia de Champions League del 19/02/2020. Asimismo, hay datos de que algunos casos importados de Italia, si bien entraron por Valencia, estuvieron en sitios públicos en Madrid. También se celebró una feria de moda en Milan casi al mismo tiempo y poco después una de arte en Madrid. Todo esto hace muy probable que el SEC8 se introdujera simultáneamente por Valencia, Madrid, y posiblemente otras ciudades y CCAA. Estos eventos concuerdan con las fechas probables de entrada de la Tabla 1. Además, el análisis (mapas en Figura 16) indica focos tempranos en el País Vasco y en Andalucía, aunque no podemos concretar si fueron introducciones independientes, posiblemente desde las mismas regiones italianas, o movimientos del virus desde Valencia o Madrid. Esto indica que probablemente el éxito inicial dependió de múltiples introducciones y no solo de una.

Sin embargo, el SEC8 en muy poco tiempo dominó la epidemia en España. La filogenia del SEC8 da indicaciones que probablemente varios eventos de superdispersión local junto con la movilidad posterior generaron la base para el éxito del SEC. Si bien, esos eventos son difíciles de definir solo a partir de la filogenia su papel queda claro cuando lo juntamos con datos epidemiológicos. El 23 de Febrero de 2020 un funeral en Vitoria con asistentes sobretodo del País Vasco y de La Rioja originó un evento superdispersor. Dicho evento pertenece al SEC8 y los casos se encuentran en la base del SEC8, al igual que los de los importados desde Italia (Figura 16). A partir de ahí el virus se expandió desde Madrid, Valencia y País Vasco/La Rioja al resto del país, sobretodo asociado a lo que creemos que son eventos locales de superdispersión (marcados con una flecha azul en Figura 16).



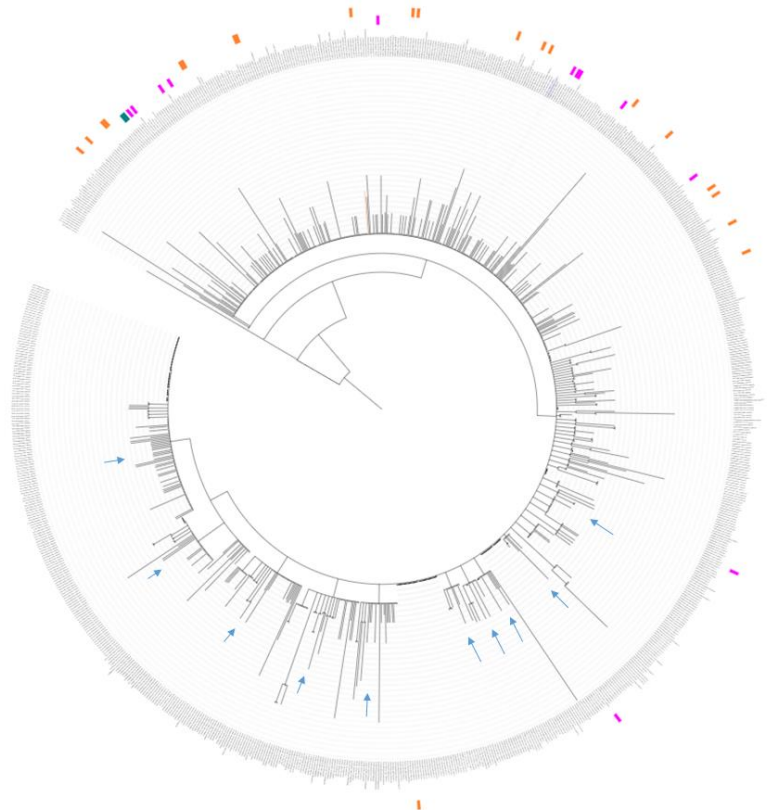
MINISTERIO
DE CIENCIA
E INNOVACIÓN



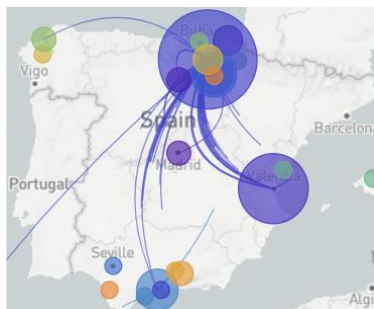
ibv INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

Tree scale: 0.0001

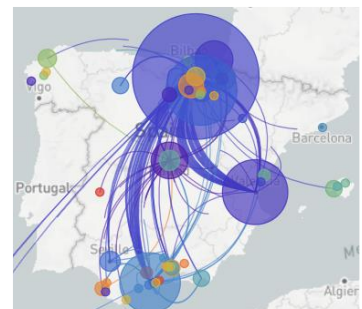
- Italy - Valencia
- Italy - Madrid
- Burial superspreader event
- Likely suprespreader events



20/02/2020



29/02/2020



14/03/2020

Figura 16. Eventos relacionados con el éxito epidemiológico del SEC8. La filogenia contempla todas las muestras asignadas a SEC8 en nuestros análisis. En diferentes colores se identifican casos de los que tenemos constancia epidemiológica de sus movimientos. Dos casos están asociados a una convención en Italia y entraron en Valencia aunque estuvieron en Madrid durante el periodo infeccioso. Varios casos fueron contagiados en el partido de fútbol celebrado entre el Bérnago y el Valencia. Además, en la base de la filogenia también se identifican casos



asociados a un funeral en Vitoria. Dichos casos generaron un gran número de eventos secundarios como los identificados con color azul que después se extendieron a una residencia de ancianos y a la población general. Mapa del movimiento predicho de las cepas pertenecientes al SEC8 de acuerdo a NextSpain (<http://seqcovid.csic.es/nextspain/>).

9. Datos preliminares de segunda ola

En la segunda ola identificamos una presencia muy residual de los SEC. Prácticamente todos se han extinguido, indicando el gran éxito que representó el confinamiento como se puede observar en Figura 14.

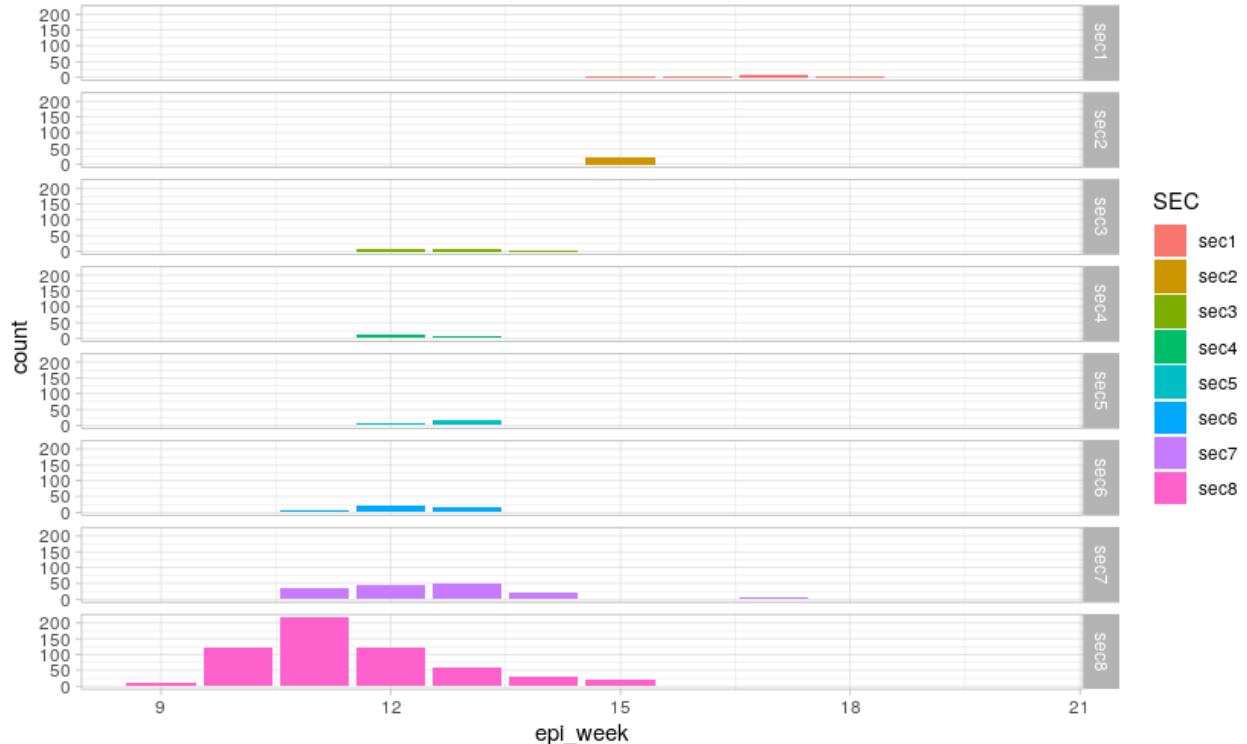


Figura 14. Número de casos asignados a cada SEC por semana epidémica.

Sin embargo, observamos la expansión de otros genotipos reproduciendo patrones parecidos que llevaron al éxito del SEC8. Tenemos datos preliminares que nuevos SEC han reemplazado a los de la primera ola, pero siguiendo un mismo patrón. Esperamos dar a conocer esos resultados muy pronto.



MINISTERIO
DE CIENCIA
E INNOVACIÓN



ibv INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

10. Epidemiología genómica para identificación y confirmación de brotes epidemiológicos

El virus SARS-CoV-2 evoluciona más lentamente que otros virus conocidos como el de la gripe o el del VIH. La tasa de evolución se sitúa en aproximadamente dos cambios nucleotídicos cada mes. Dicho de otra manera, un virus secuenciado en Septiembre de 2020 estará a alrededor de 15-20 mutaciones de distancia con respecto a los primeros virus secuenciados en la pandemia. Esta tasa de mutación limitada impacta, por ejemplo, en nuestra capacidad de usar solamente la secuencia genómica para delimitar un grupo de transmisión. Aún así, hemos estado usando la información genómica para ayudar en las investigaciones epidemiológicas en diferentes escenarios. El procedimiento a seguir se describe en la siguiente figura, ilustrando con un ejemplo real de petición del Hospital Universitario de Donostia.

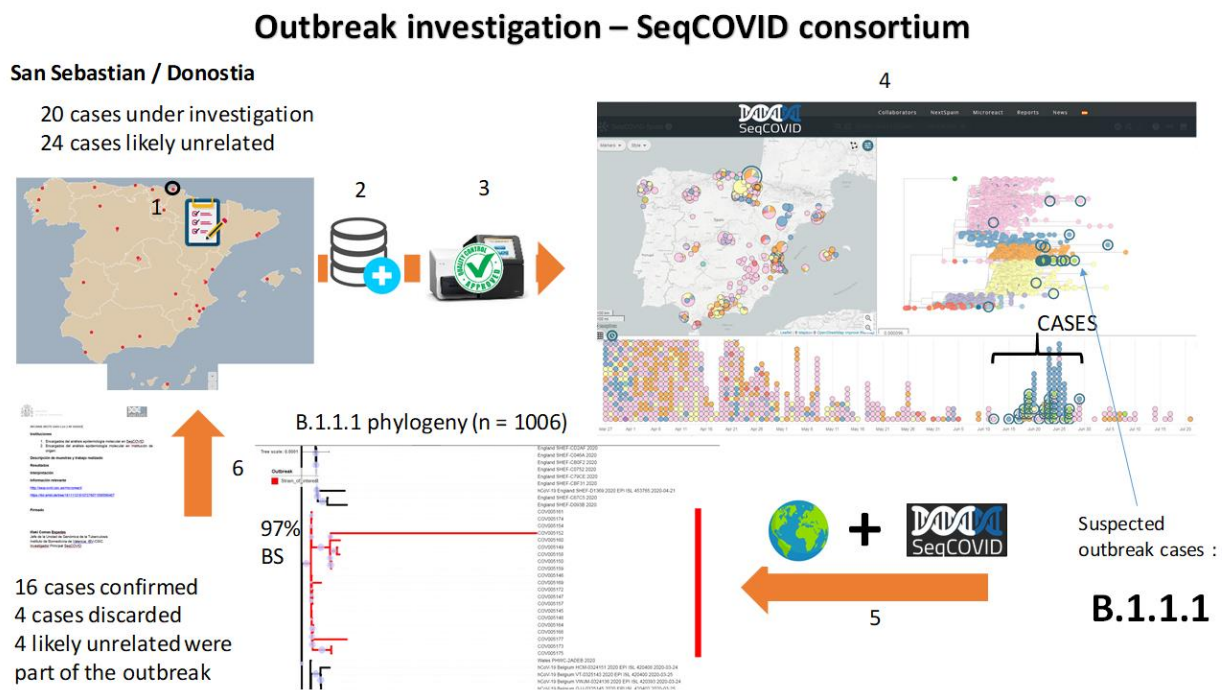


Figure 15. Los pasos para la investigación de brotes seguidos en el consorcio SeqCOVID. 1) Se recibe información de posible brote entre las muestras que se remiten del hospital al centro de secuenciación; 2) la información se graba, las muestras se secuencian y las que 3) pasan el control de calidad se analizan; 4) se analiza el brote en el marco del resto de secuencias de España. Un requisito necesario, aunque no suficiente, es que agrupen en la filogenia y sean genéticamente cercanas; 5) se reconstruye una filogenia más detallada usando las muestras sospechas corroboradas en el paso 3 más todas las muestras del linaje y todas las muestras



MINISTERIO
DE CIENCIA
E INNOVACIÓN



INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

disponibles hasta ese momento de otras partes del mundo; se identifica cualquier clado específico de brote con soporte de *bootstrap* al menos 80 o definido por 1 SNP propio; 6) se hace un informe sobre cada muestra secuenciada y si pertenece o no al brote.

En el ejemplo que ilustramos podemos observar en la siguiente tabla que la secuenciación genómica ayuda a incluir y excluir casos de un brote así como a agrupar lo que originalmente eran brotes separados en un solo brote:



ID	Sospecha hospital	Resultado_genomas_brotes
COV005136	pertenece a brote	No pertenece a brote
COV005137	caso familiar 1	brote2 confirmado
COV005138	pertenece a brote	calidad insuficiente
COV005139		brote3 confirmado
COV005140		No pertenece a brote
COV005141		No secuenciado
COV005142		brote2 confirmado
COV005143		brote3 confirmado
COV005144		No pertenece a brote
COV005145	pertenece a brote	brote1 confirmado
COV005146	pertenece a brote	brote1 confirmado
COV005147	pertenece a brote	brote1 confirmado
COV005148	pertenece a brote	brote1 confirmado
COV005149	pertenece a brote	brote1 confirmado
COV005150	pertenece a brote	brote1 confirmado
COV005151		No pertenece a brote
COV005152	pertenece a brote	brote1 confirmado
COV005153	caso familiar 1	brote2 confirmado
COV005154	pertenece a brote	brote1 confirmado
COV005155		No pertenece a brote
COV005156		No pertenece a brote
COV005157	pertenece a brote	brote1 confirmado
COV005158	pertenece a brote	brote1 confirmado
COV005159	pertenece a brote	brote1 confirmado
COV005160	pertenece a brote	brote1 confirmado
COV005161	pertenece a brote	brote1 confirmado
COV005162	caso familiar 2	brote4 confirmado
COV005163		brote2 confirmado
COV005164	pertenece a brote	brote1 confirmado
COV005165	caso familiar 2	brote4 confirmado
COV005166	caso familiar 2	brote4 confirmado
COV005167	pertenece a brote	No secuenciado
COV005168	pertenece a brote	brote1 confirmado
COV005169	pertenece a brote	brote1 confirmado
COV005170		No secuenciado
COV005171		calidad insuficiente
COV005172	caso familiar 3	brote1 confirmado
COV005173	caso familiar 3	brote1 confirmado
COV005174	caso familiar 3	brote1 confirmado
COV005175	caso familiar 3	brote1 confirmado
COV005176		calidad insuficiente
COV005177		calidad insuficiente
COV005178	¿Pertenece a brote?	No pertenece a brote

Tabla 2. Resultados de análisis de varios brotes en investigación en la región sanitaria asociada al Hospital Universitario de Donostia.



Sin embargo, hay que tener en cuenta que la tasa de evolución del coronavirus es baja, de alrededor dos mutaciones al mes. Este hecho limita la resolución de los análisis filogenéticos en muchos casos, haciendo difícil distinguir las muestras de un brote de otras muestras filogenéticamente relacionadas. Por tanto, la epidemiología genómica aquí juega un papel complementario al de la investigación epidemiológica convencional.

11. Limitaciones del consorcio

SeqCOVID está representando una experiencia única en la introducción de la epidemiología genómica a escala nacional en España pero hay toda una serie de problemas a los que nos enfrentamos y que nos impide un progreso como el que querríamos. Particularmente la epidemiología genómica tiene gran potencial cuando se puede aplicar en tiempo real de tal manera que se puede hacer vigilancia epidemiológica y ayudar en el manejo de la pandemia en el momento que está ocurriendo.

1- Acceso a muestras. En época de pandemia el sistema nacional de salud está sobrecargado por lo que las muestras no pueden ser remitidas para su análisis en tiempo real. No existen mecanismos que alivien la carga en el hospital de preparar muestras y datos asociados.

2- No existe una red de interacción entre los hospitales y los centros de secuenciación para recoger el porcentaje mayor de muestras posibles. Aún recogiendo más de 30 hospitales nuestro consorcio no ha incluido muchos hospitales, especialmente aquellos de rango menor que, sin embargo, son importantes para entender las epidemias locales. En algunas comunidades se ha solucionado aliándonos con consorcios locales que sí tienen acceso a esas muestras.

3- Acceso a datos. Los datos de los pacientes (clínicos, epidemiológicos) no se disponen en las unidades de microbiología clínica o el acceso a los mismos es muy difícil, tanto por razones logísticas como por impedimentos legales o de autorización previa. Por ello, para la mayoría de casos de SeqCOVID solo tenemos un número limitado de variables recogidas. Ligar esos casos a datos epidemiológicos es casi imposible dada la gran variedad de instituciones que deben involucrarse para su obtención.

4- La información en tiempo real, como ocurre en el Reino Unido, requiere de personal disponible que esté dedicado a tiempo completo a generar informes. SeqCOVID no dispone de ese personal por falta de presupuesto aún cuando sí se ha incorporado personal para las áreas de secuenciación y logística. El tiempo real también requiere de una disponibilidad rápida de muestras.



Conclusiones

Aún con las limitaciones de muestreo y a que el análisis es retrospectivo y no en tiempo real, el consorcio ha sido capaz de analizar los eventos ocurridos en la primera ola de la COVID-19. Más allá de la investigación académica creemos que nuestros resultados permiten extraer lecciones para el control de la pandemia:

1. La estructura poblacional del virus en España es muy diferente a la de restos de países de nuestro entorno estando enriquecida en linajes basales, del principio de la epidemia en Asia. Creemos que debe reflejar la estructura en Italia pero la falta de secuencias de ese país hace que no podamos comprobarlo.
2. Unos pocos casos ya son indicadores de una situación que se puede descontrolar. En Febrero el número de casos de SEC8 era apenas unas docenas y sin embargo terminó representando un 30% de la primera ola.
3. El cierre de fronteras a países con alta incidencia es relevante para evitar la importación simultánea y múltiple del virus. Los SEC más exitosos fueron aquellos que llegaron pronto y por múltiples vías. El cierre tardío no impidió la expansión interna de esos linajes.
4. Las ciudades actúan de irradiadora del virus. El control de la movilidad es fundamental cuando el nivel de transmisión comunitaria es alto.
5. El SEC8 es un buen ejemplo de por qué la epidemia enraíza más fuertemente en algunos lugares que en otros. En este caso, en España entró simultáneamente desde Italia por diferentes vías, probablemente no todas documentadas por nosotros. Algunas de esas entradas murieron, pero otras lograron establecerse y expandirse primero a partir de eventos de superdispersión locales (similar a lo descrito en los Estados Unidos) y luego a través de la movilidad a través del país. Como resultado, a principios de Marzo de 2020 el SEC8 dominaba la epidemia en España y se encontraba más o menos presente en todo el territorio nacional. Un control de fronteras más estricto unido a cierres de movilidad tempranos y locales probablemente hubiera limitado la expansión del SEC8 y, por tanto, de la epidemia en España.
6. La epidemiología genómica en SARS-CoV-2, si bien complementa muy bien la investigación epidemiológica, tiene limitaciones. La primera y principal para entender la pandemia en España es la falta de muestreo genómico en Italia, país directamente relacionado con nuestra pandemia. Eso hace que no podamos encontrar muchas veces la relación entre los virus de España y los de Italia. Sin embargo, la información de las encuestas epidemiológicas ha sido crucial para suplir las limitaciones de muestreo. Por ejemplo, en el SEC8 la información epidemiológica muestra como el genotipo ha sido claramente importado de Italia.
7. Esto nos lleva a la segunda limitación, la epidemiología genómica es más potente cuando más información epidemiológica y clínica tenga asociada. A día de hoy ha sido muy difícil ligar los casos secuenciados con la información individual disponible en otras bases de

datos. Debemos por tanto avanzar hacia una estructura de datos digitalizada, centralizada y accesible que permita facilitar el análisis de cohortes como la presentada en este estudio.

8. Por último, la vigilancia en tiempo real, el objetivo al que nos deberíamos dirigir, debe integrarse como parte del sistema de vigilancia española y no depender de proyectos de investigación concretos ni de la buena voluntad y esfuerzo de las unidades de microbiología clínica.

Agradecimientos

Este proyecto está financiado por el Instituto de Salud Carlos III con código COV20/00140 financiado por el Gobierno de España así como por la PTI Salud Global del Consejo Superior de Investigaciones Científicas financiada a través de donaciones de diferentes empresas que se pueden consultar (<https://pti-saludglobal-covid19.corp.csic.es/>). Para la Comunidad Valenciana hemos contado con financiación y apoyo de la Conselleria de Sanitat Universal i Salut Pública de la Generalitat Valenciana (a Fernando González Candelas). Queremos reconocer la labor de los miembros del consorcio y sus instituciones que han apoyado logísticamente y con recursos humanos el proyecto. Queremos reconocer a los pacientes participantes.

Una gran parte de los análisis presentados han sido gracias a las instalaciones de computación asociadas a la Unidad de análisis computacional y bioestadística del Centro de Investigación Príncipe Felipe. Gran parte de la secuenciación se ha llevado a cabo en el Servicio de Secuenciación y de Bioinformática de FISABIO. Determinadas muestras requirieron el uso del laboratorio de bioseguridad P2+ y P3 de FISABIO. Todas estas instalaciones han sido financiadas en gran medida por el Fondo Europeo de Desarrollo Regional “Una manera de hacer Europa”.

También queremos reconocer la labor de otros grupos de investigación del país que han producido secuencias, algunas de las cuales hemos incorporado al estudio. De los que tenemos conocimiento: Instituto de Salud Carlos III, HU 12 de Octubre, HU La Paz, Hospital Vall d’Hebron, IrsiCaixa, Biocruces, HU Donostia, Universidad Santiago de Compostela.

Por último, queremos reconocer a todos los grupos que han depositado secuencias en GISAID de otros países y que nos han permitido entender la pandemia en España. Desde SeqCOVID fomentamos el uso libre y ético de la información generada y esperamos que la información depositada por el consorcio pueda ayudar a otras muchas investigaciones en curso.

Nota sobre la Versión

Este proyecto está bajo desarrollo. El presente informe refleja las investigaciones llevadas hasta el momento, pero podría ser susceptible de cambios en el futuro. Este informe no ha sido revisado por pares pues no es un documento de publicación científica si no un informe de situación para las autoridades sanitarias.

Firmado

REFERENCIAS

- Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole Á, Pereyaslov D; WHO European Region sequencing laboratories and GISAID EpiCoV group; WHO European Region sequencing laboratories and GISAID EpiCoV group*. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill.* 2020 Aug;25(32):2001410. doi: 10.2807/1560-7917.ES.2020.25.32.2001410.
- Gardy, Jennifer L., and Nicholas J. Loman. 2017. "Towards a Genomics-Informed, Real-Time, Global Pathogen Surveillance System." *Nature Reviews. Genetics* 19 (1): 9–20.
- Gonzalez-Reiche, Ana S., Matthew M. Hernandez, Mitchell J. Sullivan, Brianne Ciferri, Hala Alshammary, Ajay Obla, Shelcie Fabre, et al. 2020. "Introductions and Early Spread of SARS-CoV-2 in the New York City Area." *Science* 369 (6501): 297–301.
- Lemieux, Jacob, Katherine J. Siddle, Bennett M. Shaw, Christine Loreth, Stephen Schaffner, Adrienne Gladden-Young, Gordon Adams, et al. 2020. "Phylogenetic Analysis of SARS-CoV-2 in the Boston Area Highlights the Role of Recurrent Importation and Superspreading Events." *medRxiv*, August, 2020.08.23.20178236.
- Rambaut, A., Holmes, E.C., O'Toole, Á. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* (2020).
- "Report into a nosocomial outbreak of coronavirus disease 2019 (COVID-19) at Netcare St. Augustine's Hospital" n.d. Accessed October 14, 2020. <https://t.co/nwBBU6oyAv?amp=1>.
- Rivett, Lucy, Sushmita Sridhar, Dominic Sparkes, Matthew Routledge, Nick K. Jones, Sally Forrest, Jamie Young, et al. 2020. "Screening of Healthcare Workers for SARS-CoV-2 Highlights the Role of Asymptomatic Carriage in COVID-19 Transmission," May. <https://doi.org/10.7554/eLife.58728>.
- Seemann, Torsten, Courtney Lane, Norelle Sherry, Sebastian Duchene, Anders Goncalves da



MINISTERIO
DE CIENCIA
E INNOVACIÓN



INSTITUT DE
BIOMEDICINA DE
VALÈNCIA

Silva, Leon Caly, Michelle Sait, et al. n.d. "Tracking the COVID-19 Pandemic in Australia Using Genomics." <https://doi.org/10.1101/2020.05.12.20099929>.

Worobey, Michael, Jonathan Pekar, Brendan B. Larsen, Martha I. Nelson, Verity Hill, Jeffrey B. Joy, Andrew Rambaut, Marc A. Suchard, Joel O. Wertheim, and Philippe Lemey. 2020. "The Emergence of SARS-CoV-2 in Europe and North America." *Science*, September. <https://doi.org/10.1126/science.abc8169>.