Supplemental Material

Phylogeography of the SARS-CoV-2 pandemic in Spain: a story of multiple introductions, micro-geographical stratification, founder effects and super-spreaders

Alberto Gómez-Carballa, Xabier Bello, Jacobo Pardo-Seco, María Luisa Pérez del Molino, Federico Martinón-Torres, Antonio Salas

Figure S1. Accumulative counts of Spanish genomes in GISAID.

Figure S2. Features of HQ *versus* LQ datasets related to quality of the sequence data.

Figure S3. Frequency of main haplogroups by Spanish regions.

Figure S4 to S10. Lineages observed in Spanish regions. Only those with >30 genomes analyzed are shown. Large regional maps were downloaded from https://d-maps.com/ that can be used freely according to https://d-maps.com/conditions.php?lang=es. Map of a) Andalusia: https://d-maps.com/carte.php?num_car=13251&lang=es; b) Basque Country: https://d-maps.com/carte.php?num_car=13481&lang=pt; c) Catalonia: https://d-maps.com/carte.php?num_car=13853&lang=es; d) Galicia: https://d-maps.com/carte.php?num_car=14017&lang=es; e) Madrid: https://d-maps.com/carte.php?num_car=14101&lang=es; f) Navarra: https://d-maps.com/carte.php?num_car=14301&lang=es; and g) Valencian Community: https://d-maps.com/carte.php?num_car=14399&lang=es. Small Spanish maps were downloaded from Wikipedia.

Figure S11. Sampling points of interpolated maps shown in Figure 2 (main text).

Figure S12. Diversity indexes in the different Spanish regions. (A) Haplotype diversity, (B) nucleotide diversity, and (C) Accumulated sequences; note that there is discontinuity in sampling from 15 April; which however, does not have impact in our study because most of the analysis focus on the beginning of the pandemic and the time right after the lockdown on 15 March.

Figure S13. Network of all Spanish genomes.

Figure S14. Boxplot and density plot of genome lifespan of identical haplotypes in the GISAID database.

Figure S15. Interpolation map of Spanish haplogroup A2 frequencies, which roughly matches the frequency of D614G. The map on the top indicates the sampling points.

Figure S16 to S21. ML phylogenies and TMRCA for the most common Spanish clades.

Figure S1.

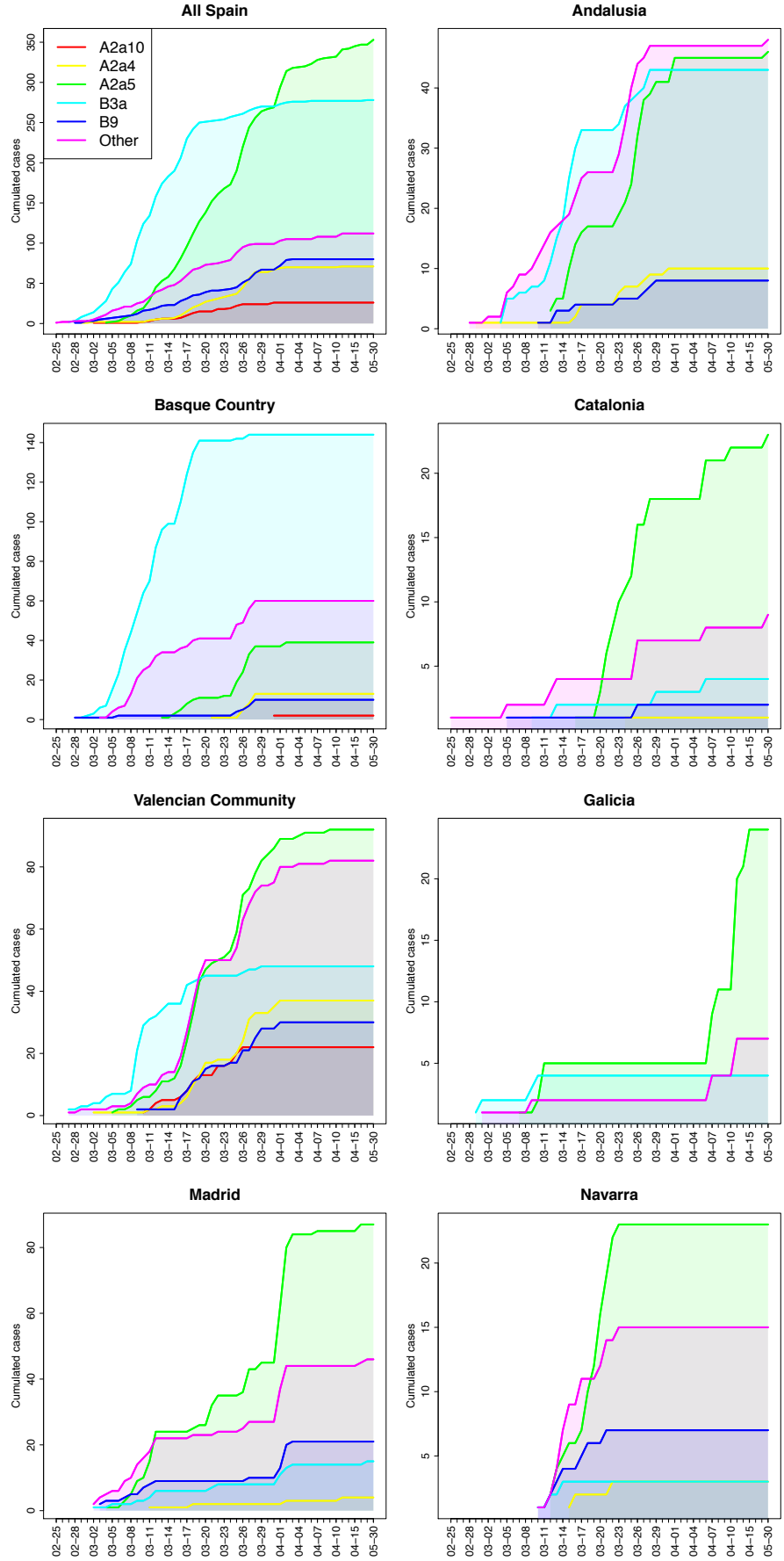

All Spain
Andalusia
Basque Country
Catalonia
Valencian Community
Galicia
Madrid
Navarra
A2a10
A2a4
A2a5
B3a
B9
Other
Cumulated cases

Figure S2.

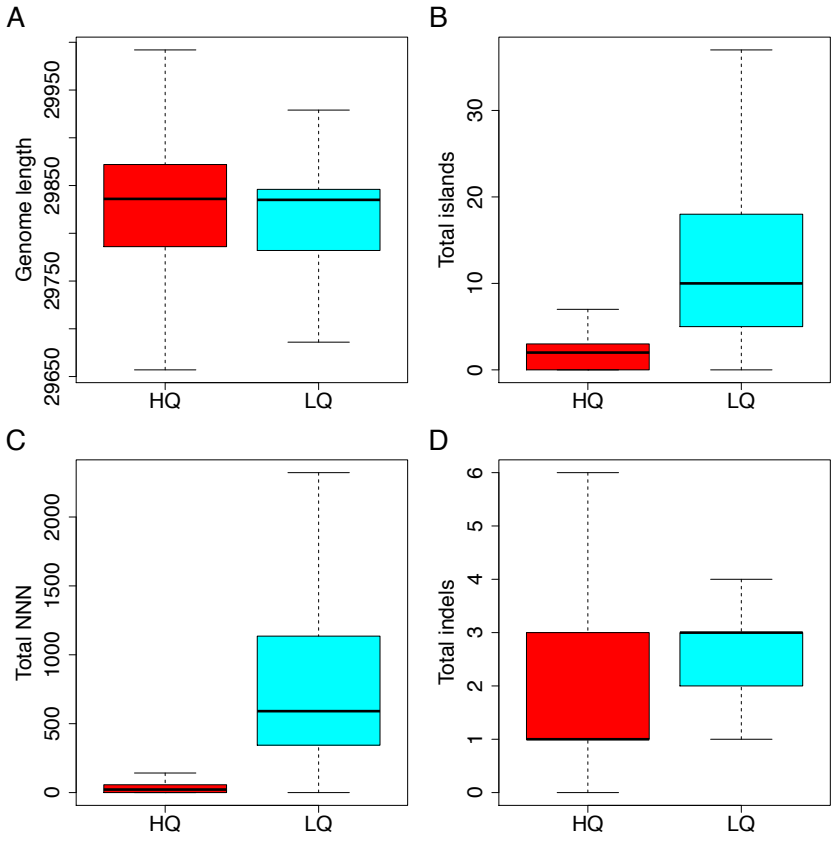


Figure S3.

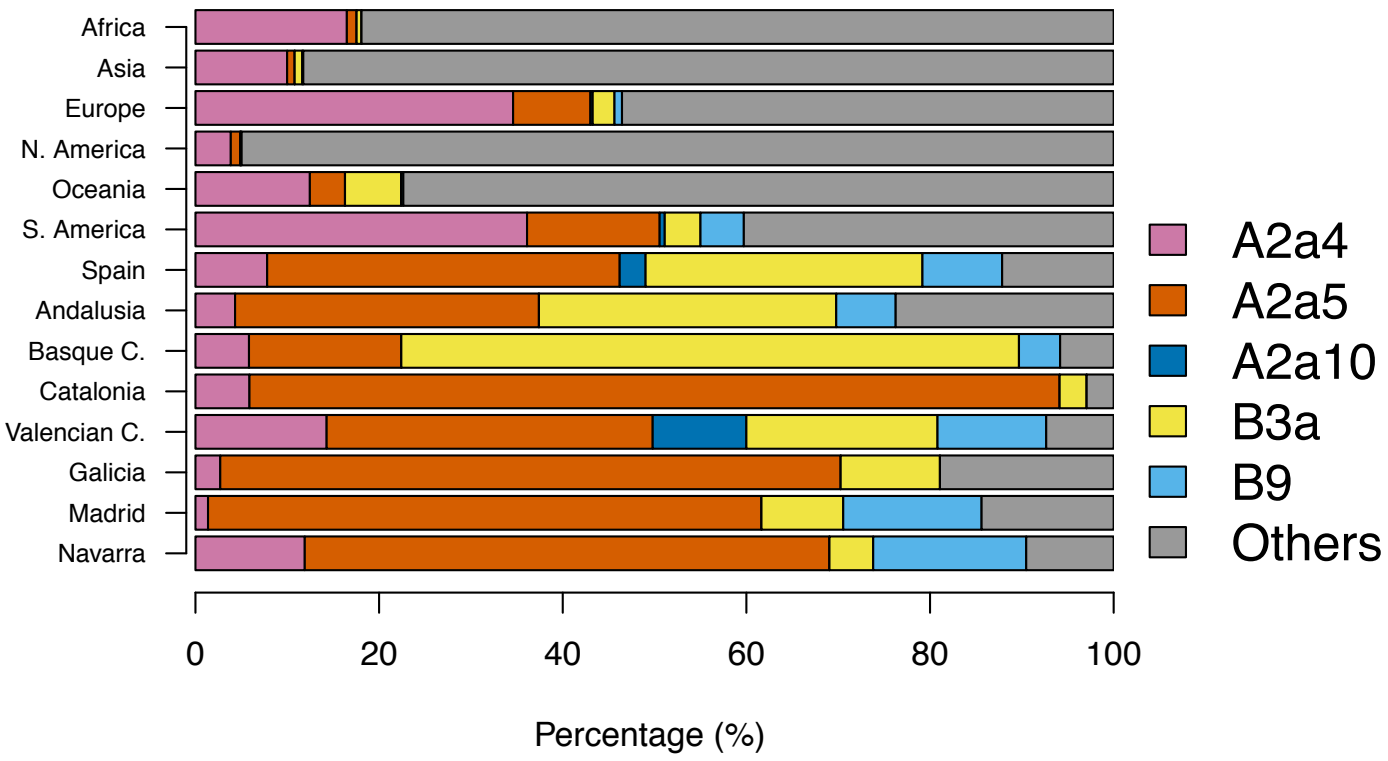


Figure S4

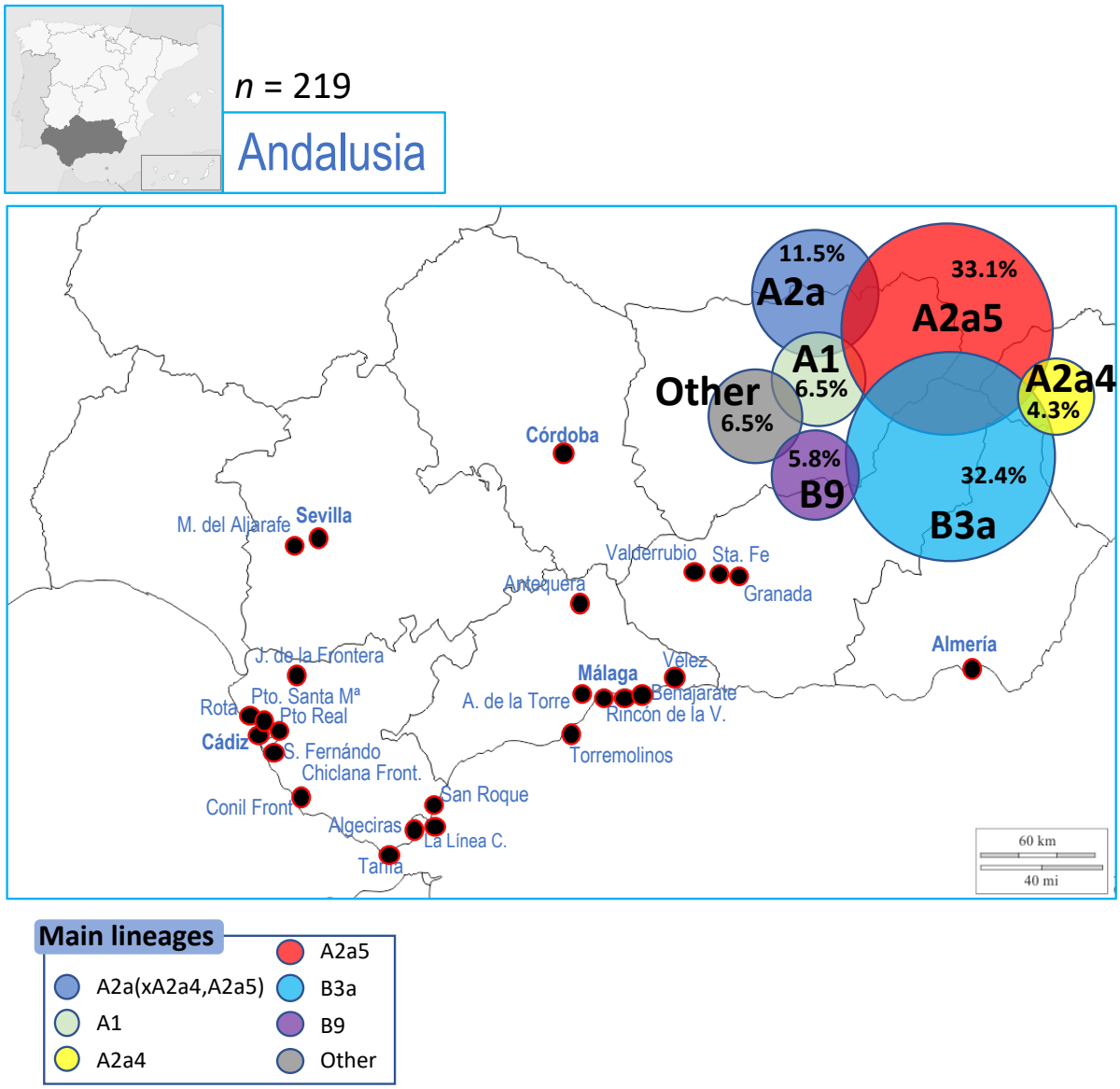

n = 219
Andalusia
11.5%
A2a
33.1%
A2a5
A1
6.5%
Other
6.5%
A2a4
4.3%
5.8%
B9
32.4%
B3a
Córdoba
M. del Aljarafe
Sevilla
Valderrubio
Sta. Fe
Granada
Antequera
Almería
J. de la Frontera
Vélez
Málaga
Rota
Pto. Santa Mª
Pto Real
A. de la Torre
Benajarafe
Rincón de la V.
Cádiz
S. Fernándo
Chiclana Front.
Torremolinos
Conil Front
San Roque
Algeciras
La Línea C.
Tarifa
60 km
40 mi
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S5

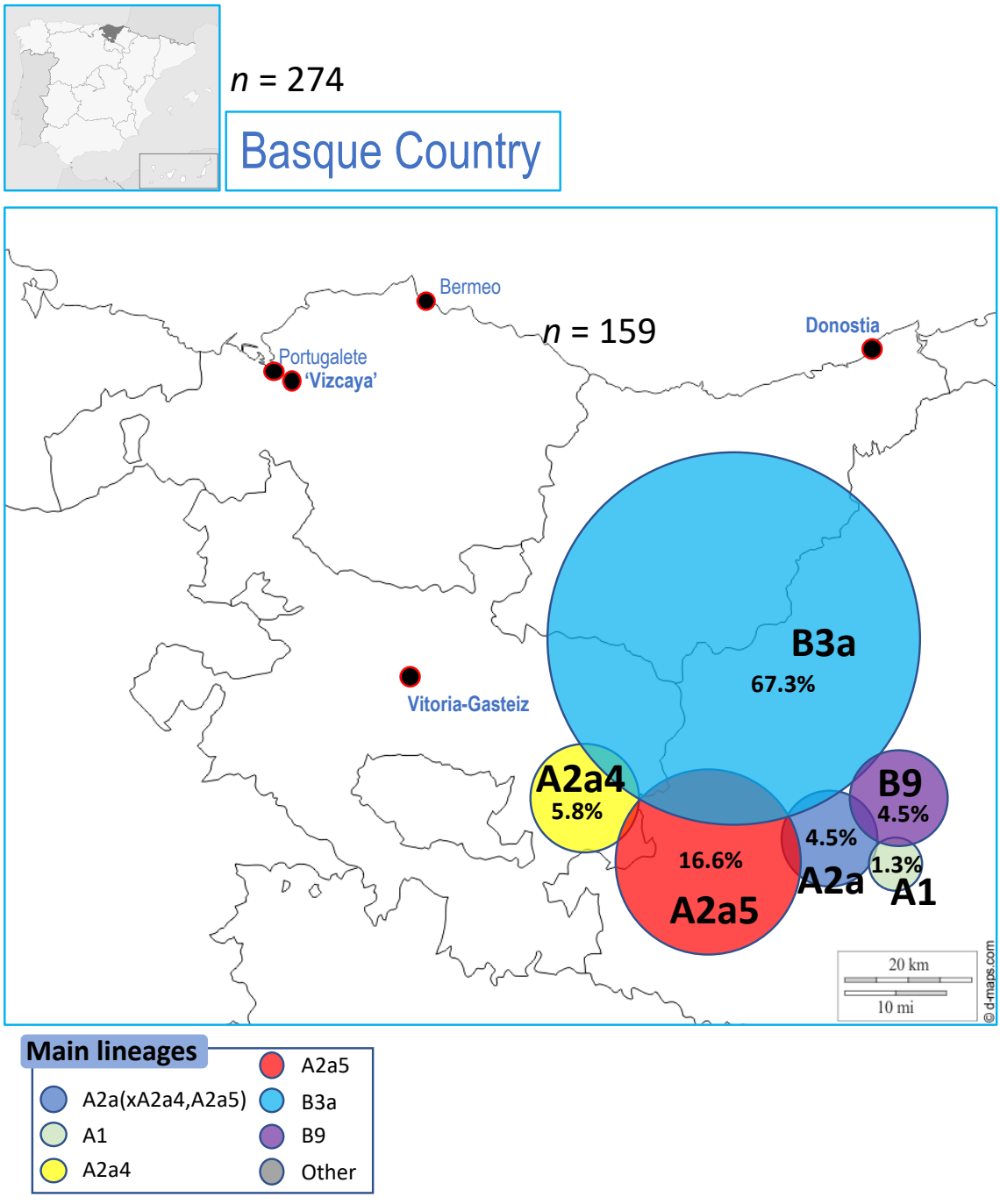

n = 274
Basque Country
n = 159
Bermeo
Portugalete
'Vizcaya'
Donostia
Vitoria-Gasteiz
B3a
67.3%
A2a4
5.8%
B9
4.5%
4.5%
16.6%
A2a5
A2a
1.3%
A1
20 km
10 mi
© d-maps.com
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S6

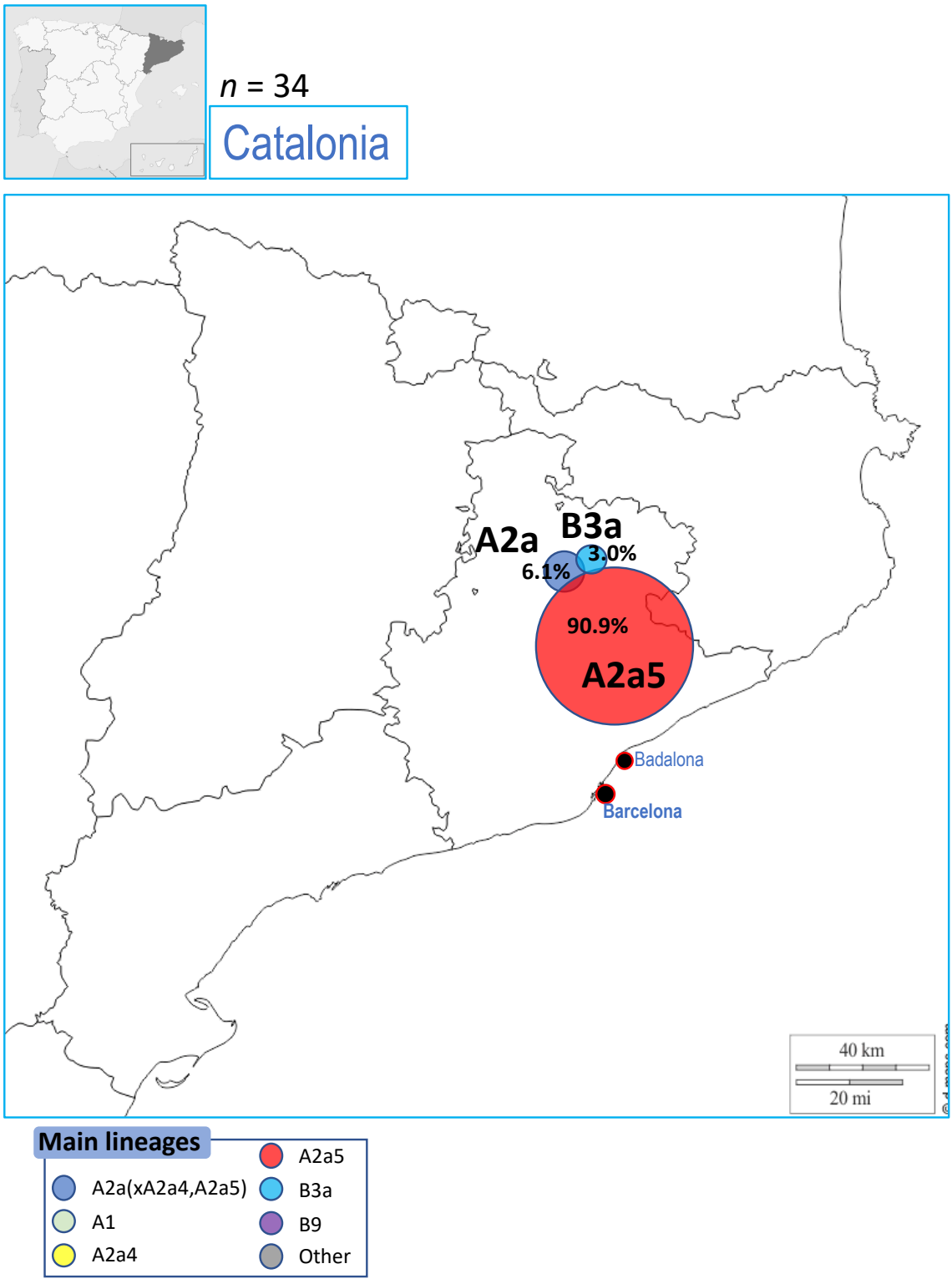

n = 34
Catalonia
A2a
6.1%
B3a
3.0%
90.9%
A2a5
Badalona
Barcelona
40 km
20 mi
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S7

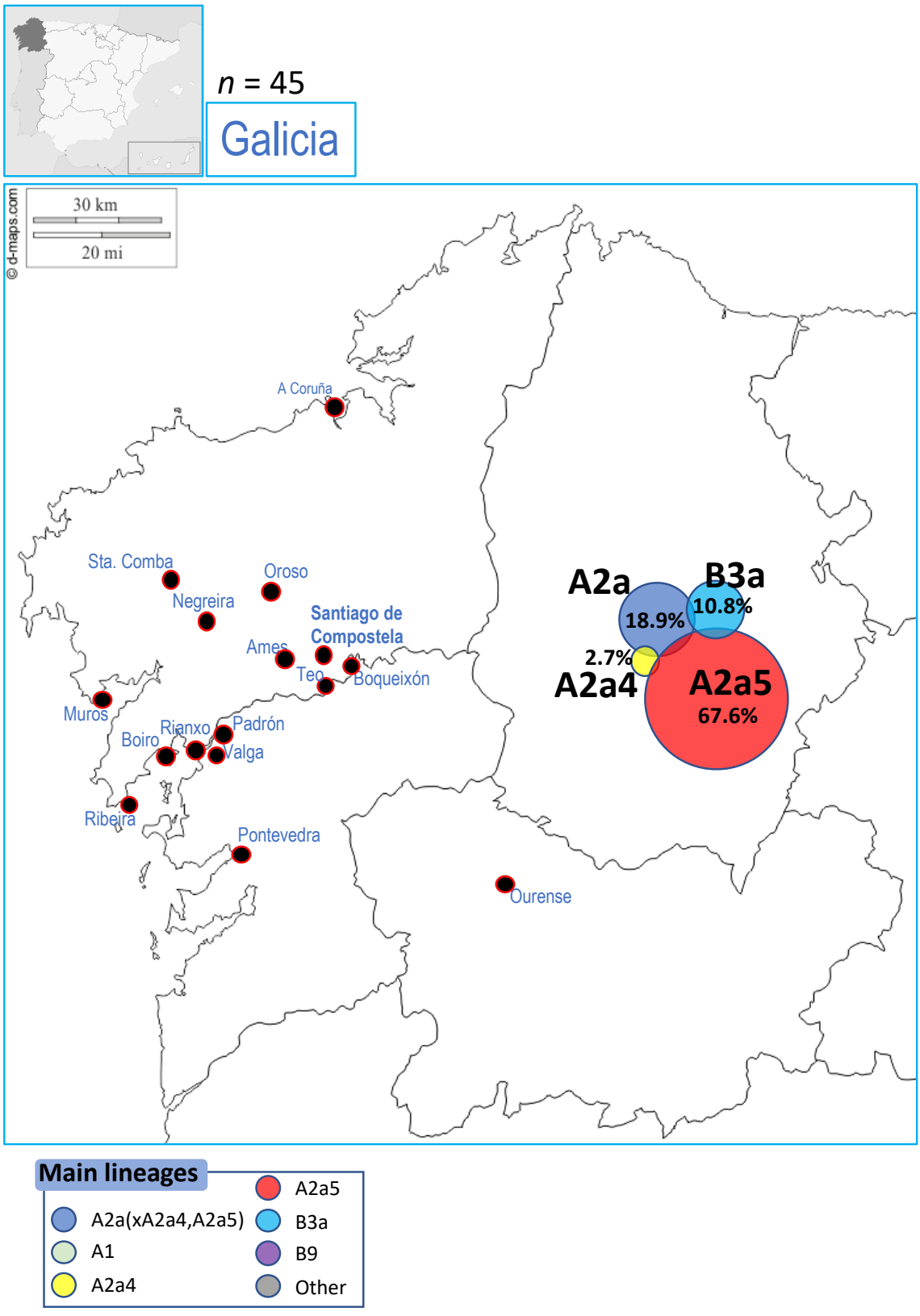

n = 45
Galicia
© d-maps.com
30 km
20 mi
A Coruña
Sta. Comba
Oroso
Negreira
Santiago de Compostela
Ames
Teo
Boqueixón
Muros
Rianxo
Padrón
Boiro
Valga
Ribeira
Pontevedra
Ourense
A2a
B3a
18.9%
10.8%
2.7%
A2a4
A2a5
67.6%
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S8

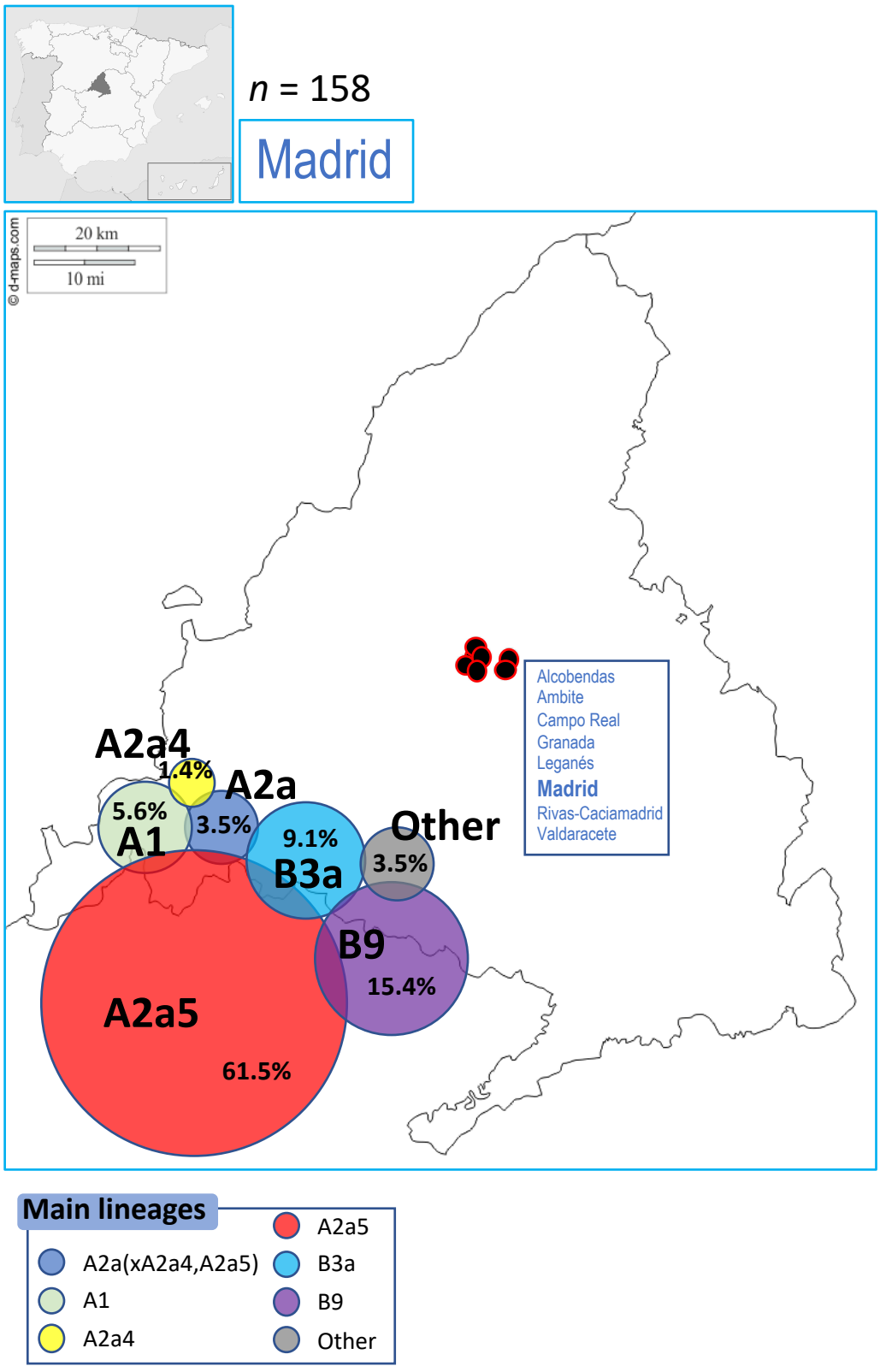

n = 158
Madrid
20 km
10 mi
© d-maps.com
Alcobendas
Ambite
Campo Real
Granada
Leganés
Madrid
Rivas-Caciamadrid
Valdaracete
A2a4
1.4%
A2a
3.5%
5.6%
A1
9.1%
B3a
Other
3.5%
B9
15.4%
A2a5
61.5%
Main lineages
A2a5
A2a(xA2a4,A2a5)
B3a
A1
B9
A2a4
Other

Figure S9

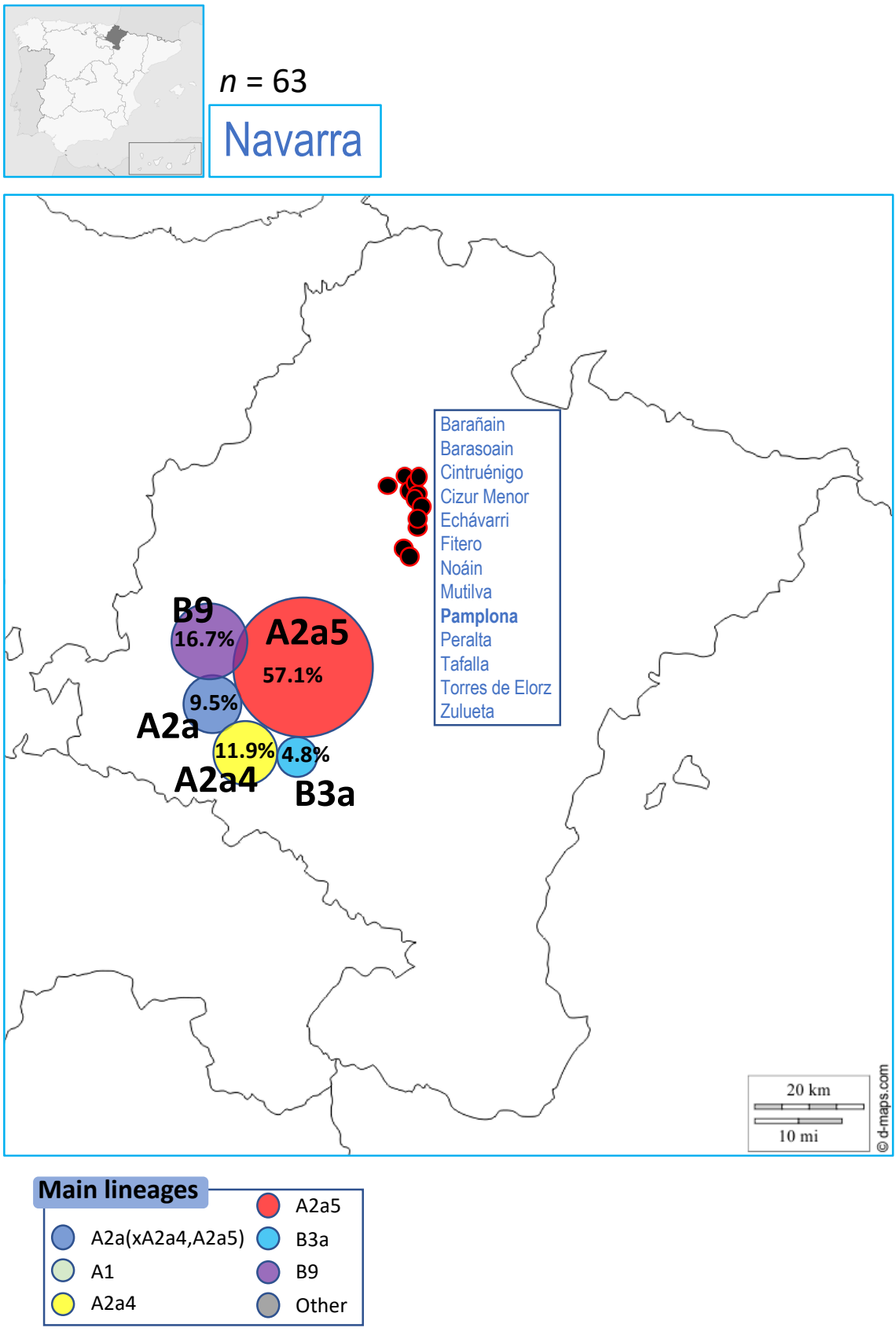

n = 63
Navarra
Barañain
Barasoain
Cintruénigo
Cizur Menor
Echávarri
Fitero
Noáin
Mutilva
Pamplona
Peralta
Tafalla
Torres de Elorz
Zulueta
B9
16.7%
A2a5
57.1%
9.5%
A2a
11.9%
A2a4
4.8%
B3a
20 km
10 mi
© d-maps.com
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S10

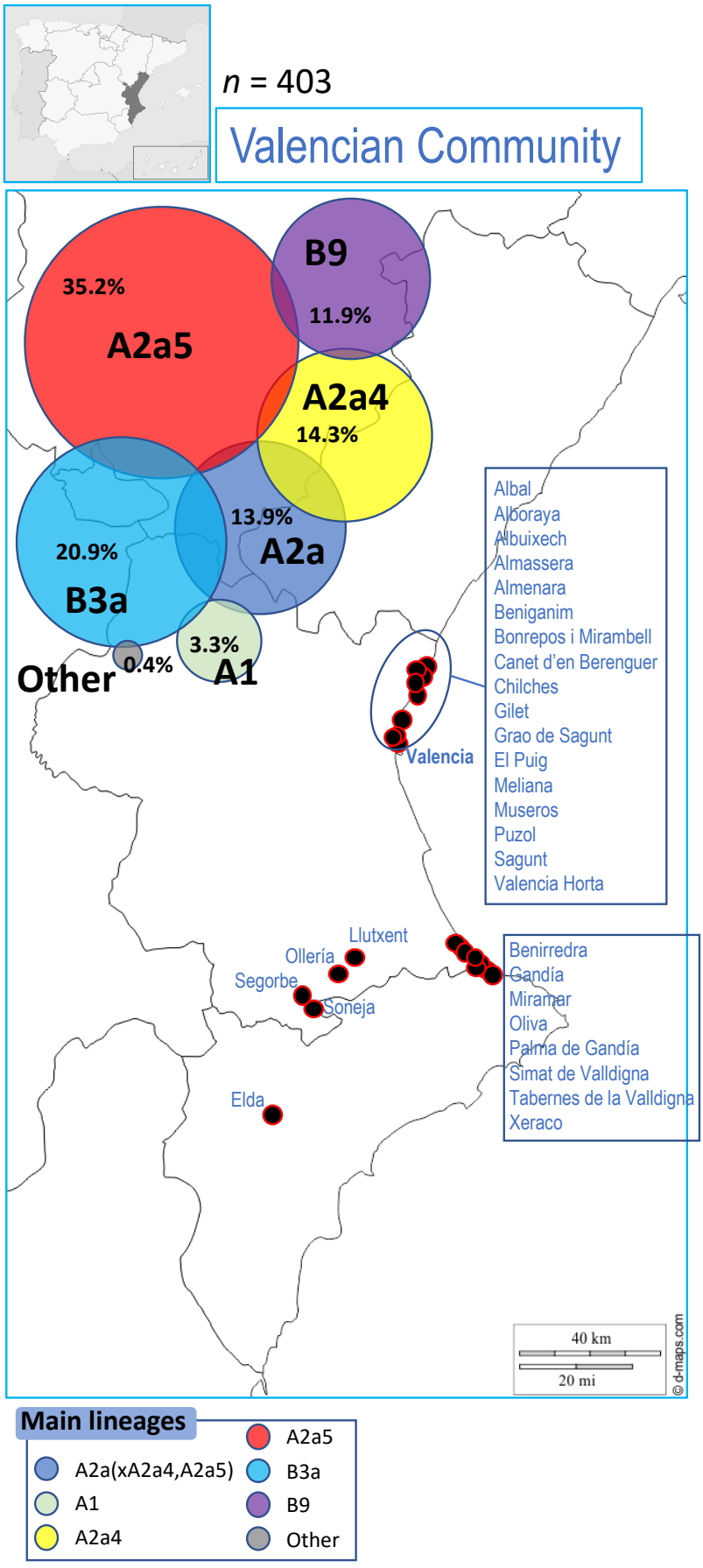

n = 403
Valencian Community
35.2%
A2a5
B9
11.9%
A2a4
14.3%
13.9%
A2a
20.9%
B3a
3.3%
A1
0.4%
Other
Valencia
Albal
Alboraya
Albuixech
Almassera
Almenara
Beniganim
Bonrepos i Mirambell
Canet d'en Berenguer
Chilches
Gilet
Grao de Sagunt
El Puig
Meliana
Museros
Puzol
Sagunt
Valencia Horta
Llutxent
Ollería
Segorbe
Soneja
Elda
Benirredra
Gandía
Miramar
Oliva
Palma de Gandía
Simat de Valldigna
Tabernes de la Valldigna
Xeraco
40 km
20 mi
© d-maps.com
Main lineages
A2a(xA2a4,A2a5)
A1
A2a4
A2a5
B3a
B9
Other

Figure S11.

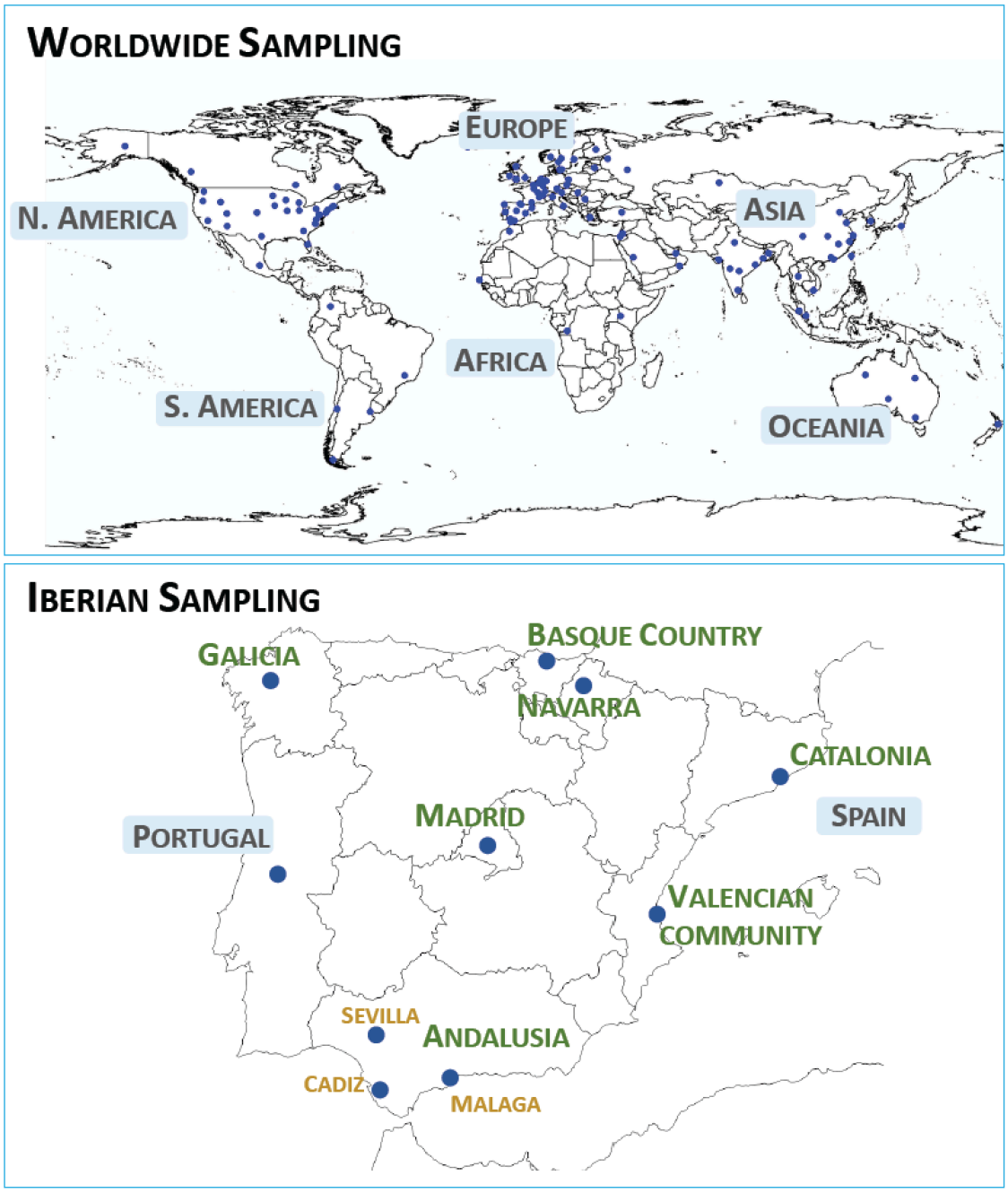

WORLDWIDE SAMPLING
EUROPE
N. AMERICA
ASIA
AFRICA
S. AMERICA
OCEANIA
IBERIAN SAMPLING
GALICIA
BASQUE COUNTRY
NAVARRA
CATALONIA
SPAIN
MADRID
PORTUGAL
VALENCIAN COMMUNITY
SEVILLA
ANDALUSIA
CADIZ
MALAGA

Figure S12.

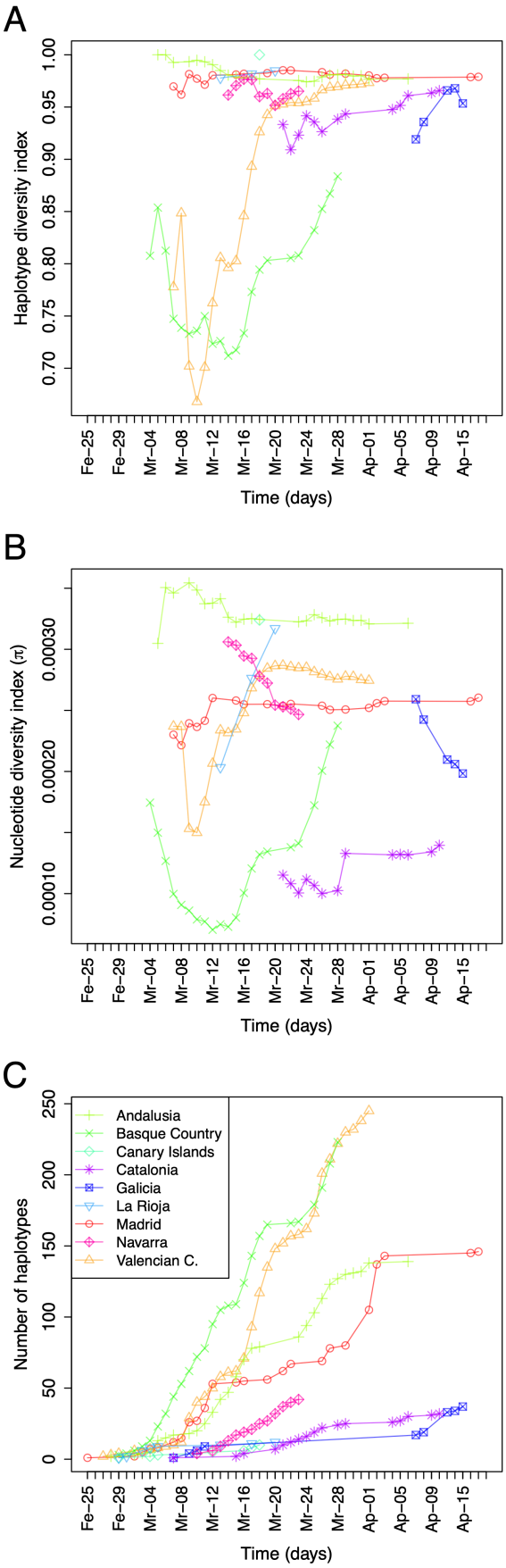

A
Haplotype diversity index
0.70
0.75
0.80
0.85
0.90
0.95
1.00
Fe–25
Fe–29
Mr–04
Mr–08
Mr–12
Mr–16
Mr–20
Mr–24
Mr–28
Ap–01
Ap–05
Ap–09
Ap–15
Time (days)
B
Nucleotide diversity index (π)
0.00010
0.00020
0.00030
Time (days)
C
Number of haplotypes
0
50
100
150
200
250
Andalusia
Basque Country
Canary Islands
Catalonia
Galicia
La Rioja
Madrid
Navarra
Valencian C.
Time (days)

Figure S13.

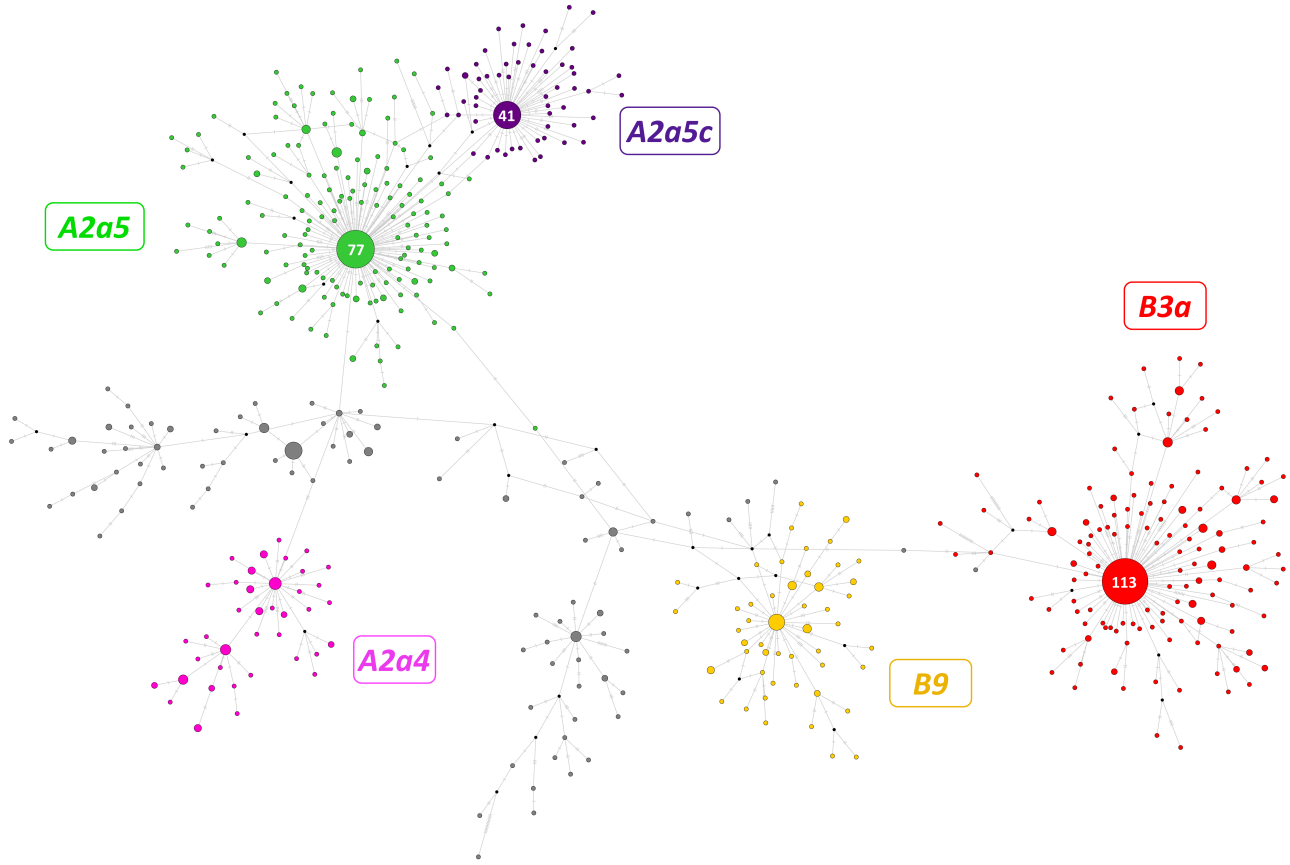


Figure S14.

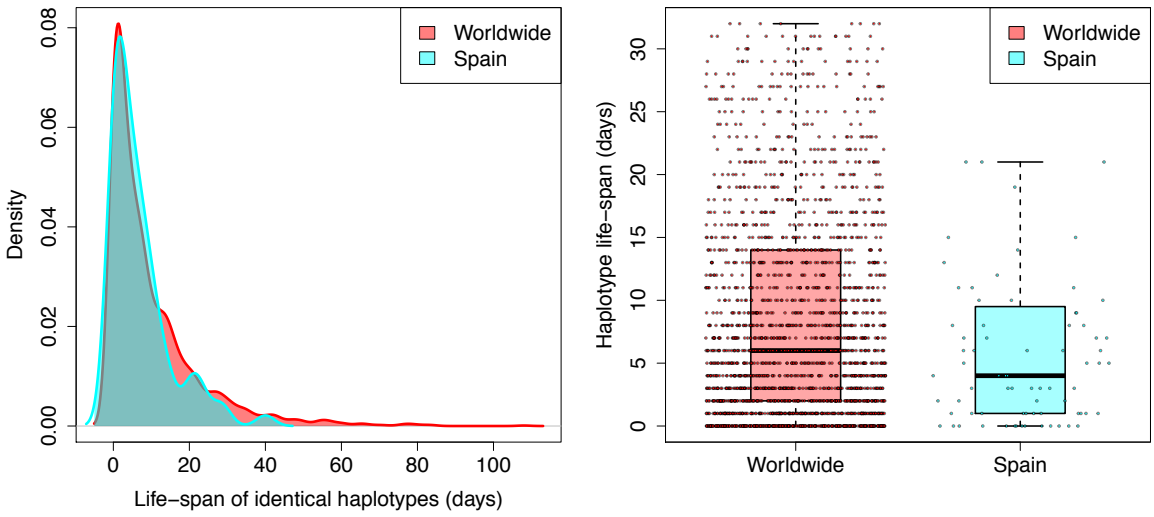
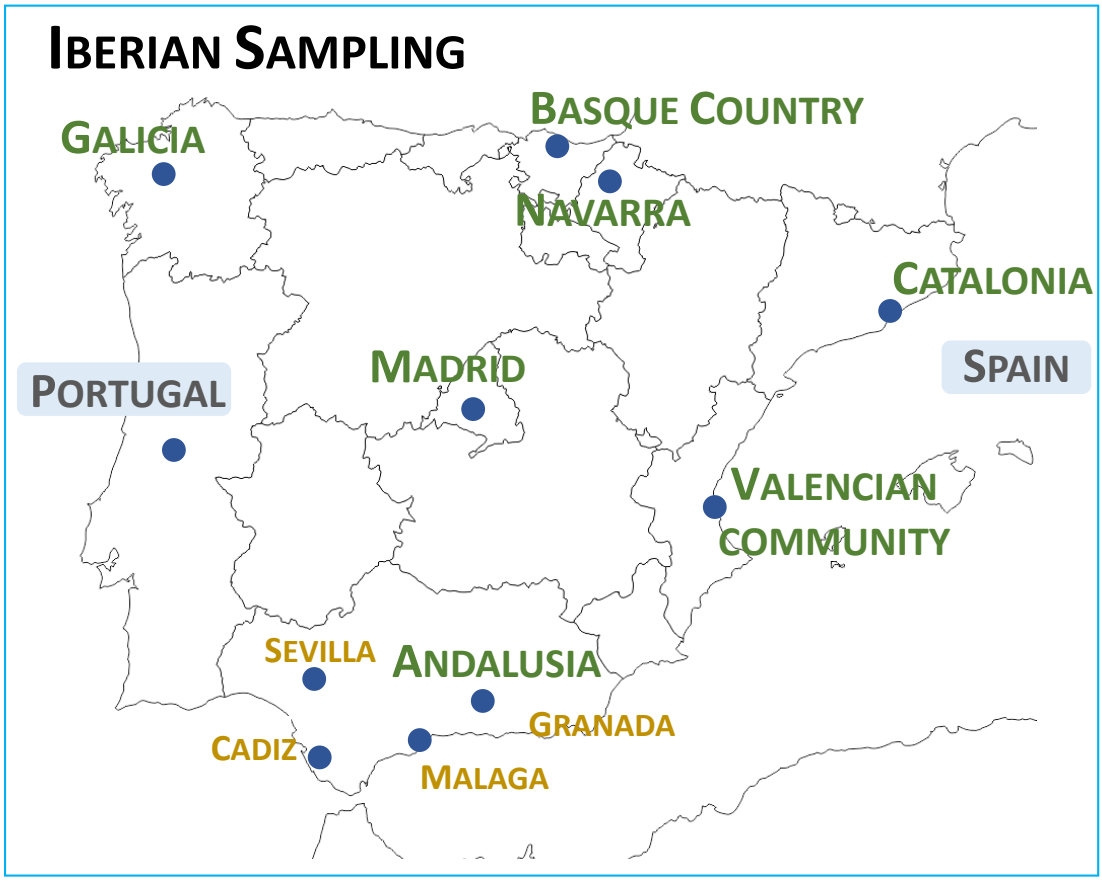


Figure S15.


IBERIAN SAMPLING
GALICIA
BASQUE COUNTRY
NAVARRA
CATALONIA
SPAIN
PORTUGAL
MADRID
VALENCIAN COMMUNITY
SEVILLA
ANDALUSIA
GRANADA
CADIZ
MALAGA

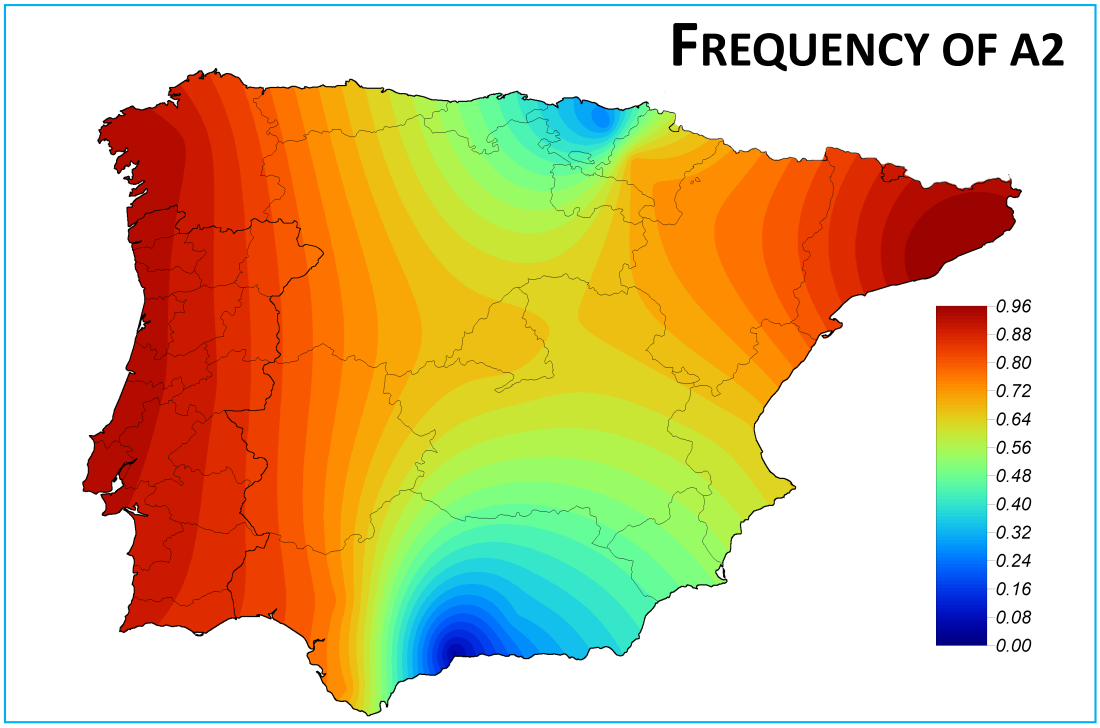

FREQUENCY OF A2
0.96
0.88
0.80
0.72
0.64
0.56
0.48
0.40
0.32
0.24
0.16
0.08
0.00

Figure S16.

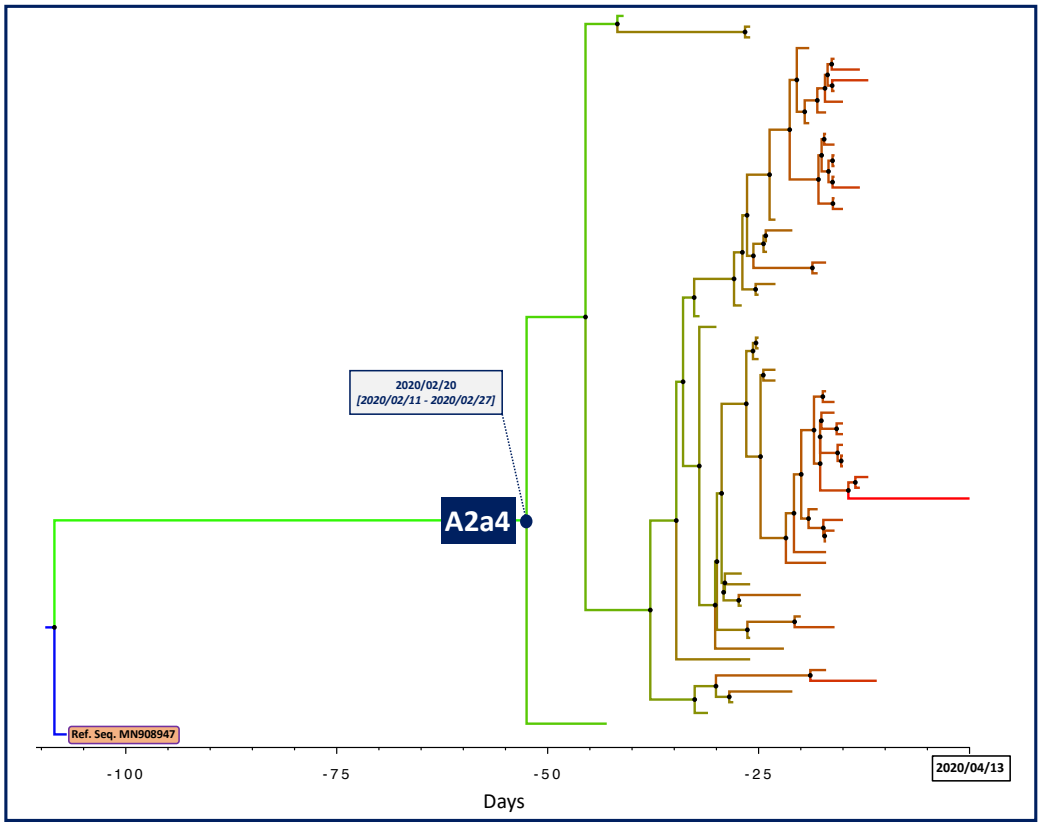


Figure S17.

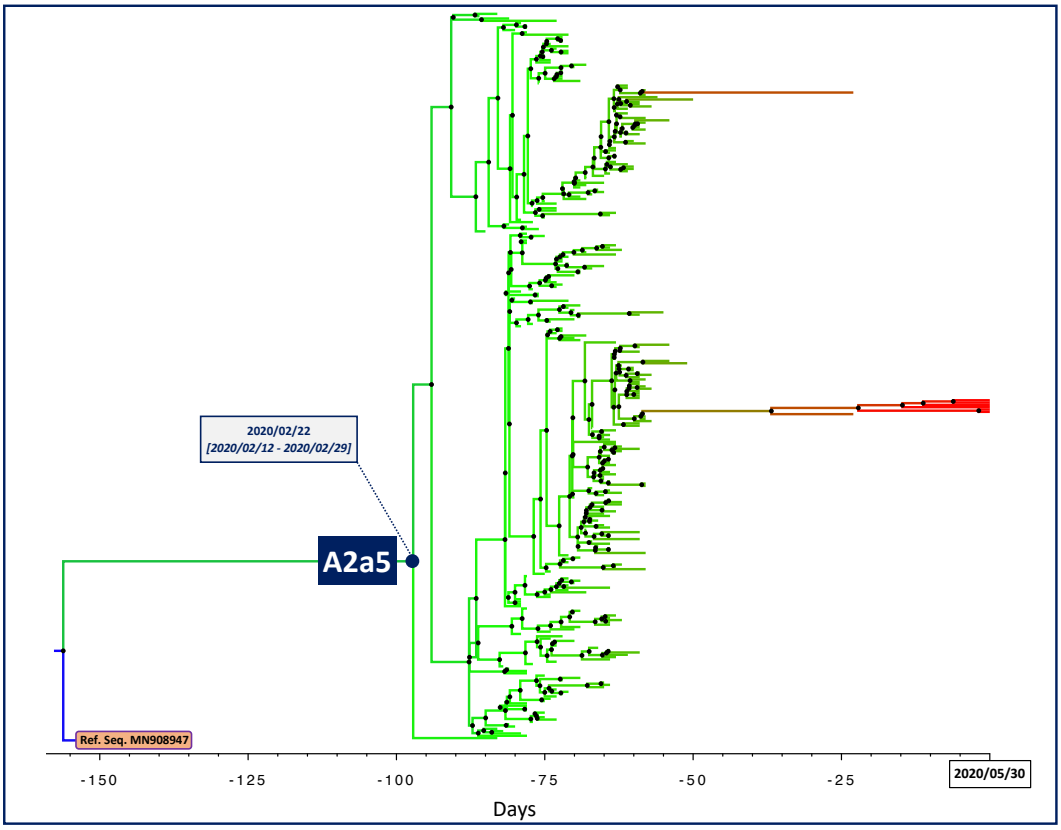


Figure S18.

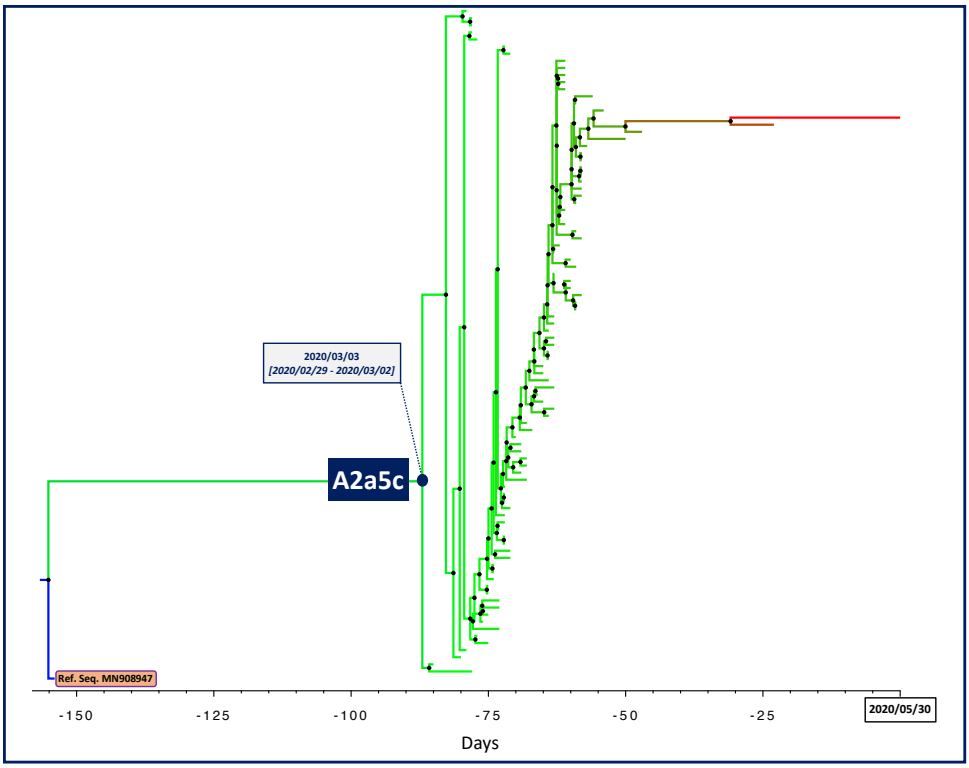


Figure S19.

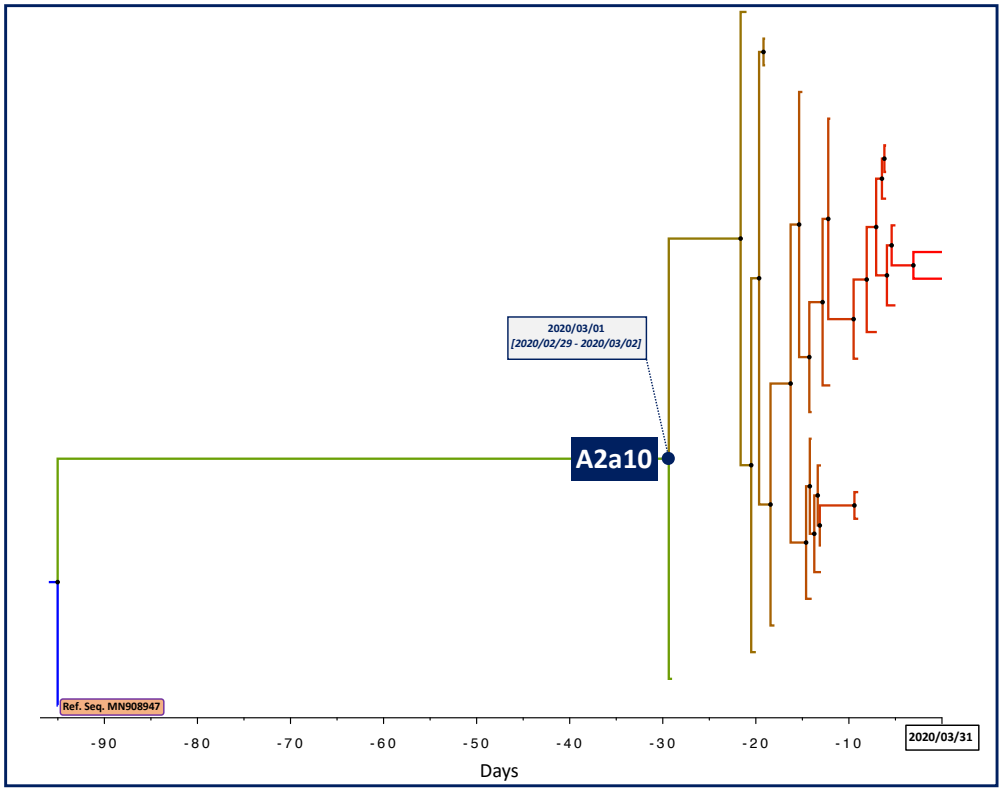


Figure S20.

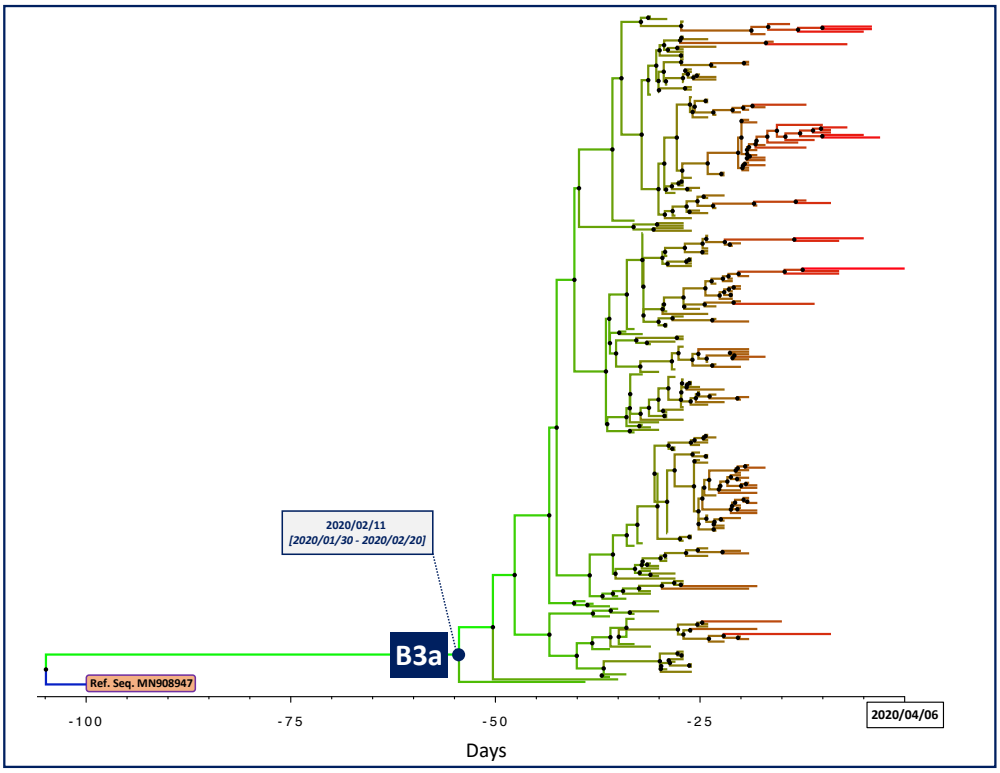
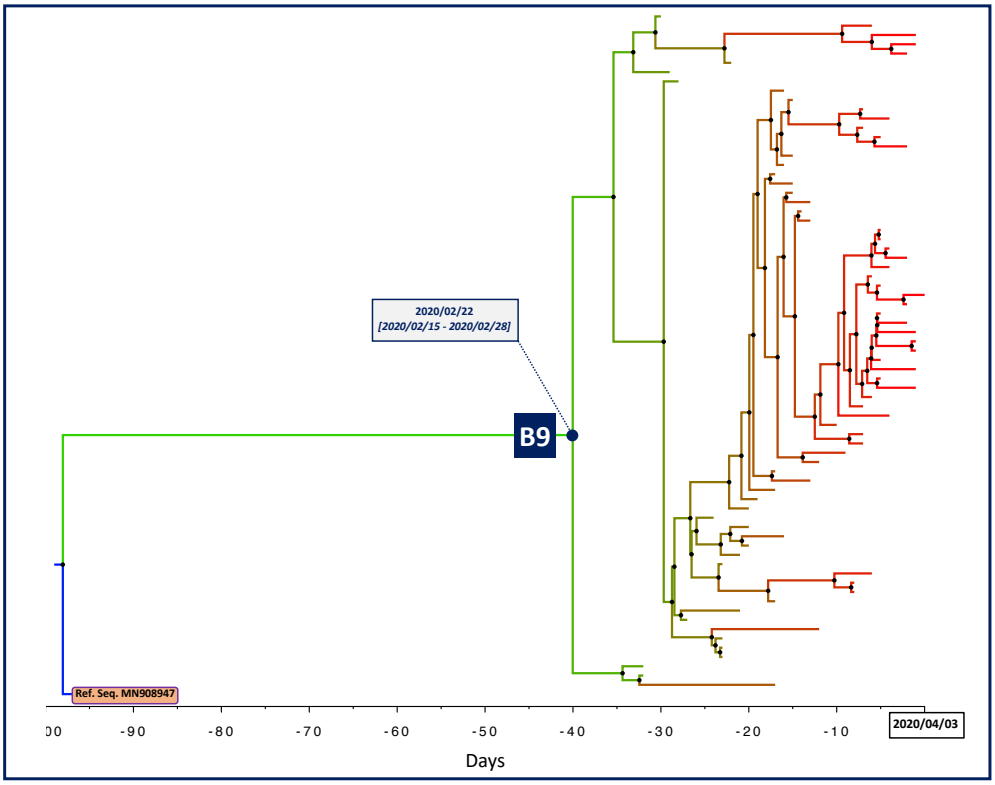



Figure S21.