

## **People from Ibiza: an unexpected isolate in the Western Mediterranean**

*Simone Andrea Biagini<sup>1</sup>, Neus Solé-Morata<sup>1</sup>, Elizabeth Matisoo-Smith<sup>2</sup>, Pierre Zalloua<sup>3</sup>, David Comas<sup>1</sup>, Francesc Calafell<sup>1</sup>*

1 Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSIC-UPF) ,  
Universitat Pompeu Fabra, Barcelona, Spain.

2 Department of Anatomy, University of Otago, Dunedin, New Zealand.

3 School of Medicine, The Lebanese American University, Chouran, Beirut, Lebanon.

Corresponding author:

Francesc Calafell

Carrer del Dr. Aiguader, 88, 08003 Barcelona, Spain

+34 933160842

[francesc.calafell@upf.edu](mailto:francesc.calafell@upf.edu)

### **Conflict of interest statement**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

## **ABSTRACT**

In this study, we seek to understand and to correlate the genetic patterns observed in the population of the island of Ibiza in the Western Mediterranean basin with past events. Genome-wide genotypes of 189 samples representing 13 of 17 regions in Spain have been analyzed, in addition to 105 samples from the Levant, 157 samples from North Africa, and one ancient sample from the Phoenician Cas Molí site in Ibiza. Before the Catalans conquered the island in 1235 CE, Ibiza (*Eivissa*) had already been influenced by several cultures, starting with the Phoenicians, then the Carthaginians, followed by the Umayyads. The impact of these various cultures on the genetic structure of the islanders is still unexplored. Our results show a clear distinction between Ibiza and the rest of Spain. To investigate whether this was due to the Phoenician colonization or to more recent events, we compared the genomes of current Ibizans to that of an ancient Phoenician sample from Ibiza and to both modern Levantine and North African genomes. We did not identify any trace of Phoenician ancestry in the current Ibizans. Interestingly, the analysis of runs of homozygosity and changes in the effective population size through time support the idea that drift has shaped the genetic structure of current Ibizans. In addition to the small carrying capacity of the island, Ibiza experienced a series of dramatic demographic changes due to several instances of famine, war, malaria and plague that could have significantly contributed to its current genetic differentiation.

## **Keywords**

Ibiza, Human Population Genetics, Genetic Isolate, Runs of homozygosity, Mediterranean

## INTRODUCTION

Island populations tend to be small and isolated, and thus prone to the effects of genetic drift, particularly through bottlenecks and founder effects. The outcome is represented by reduced genetic diversity, extended identity by descent (IBD) sharing, and a higher number of runs of homozygosity (ROHs) <sup>1</sup>. The loss of genetic diversity after a sudden decrease in population size is a process that can be more easily seen in environments that have experienced geographical and/or cultural isolation. Culturally speaking, islands represent both shelters for the conservation of traditions, and innovative places for driving processes of social change <sup>2</sup>. In the landscape of the Western Mediterranean, the complex history of Ibiza stands out for its peculiarity. Ibiza (*Eivissa* in Catalan) is one of the Balearic Islands, the closest to the continent (~90 Km away) but also the most arid. The island has an area of 577 Km<sup>2</sup>, and, until the 1960s, its population never exceeded 35,000 inhabitants, while currently it is ~140,000 (ref. 3). Archeological evidence suggests that the island was sporadically visited from the late 6th millennium to the early 5th millennium BCE but was probably not permanently settled until much later. No further evidence of human presence is found until the late 3rd millennium BCE, in the Neolithic/Calcolithic. The Bronze Age, megalithic *talaiòtic* culture of Mallorca and Minorca is absent in Ibiza, probably due to the low carrying capacity of the island resulting in a sparse population density. This also explains how the Phoenicians could settle in Ibiza with little apparent resistance <sup>4</sup>. The Phoenician *Ibossim* was founded in 654 BCE and, given the scarcity of pre-Phoenician sites discovered to date, it is suggested that the Phoenicians were the first permanent settlers of the island <sup>5</sup>. Around 550 BCE the Carthaginians, who were the descendants of the Phoenician founders of Carthage, replaced the early Phoenicians settlers <sup>6</sup> and inhabited

the island for the next five centuries, changing the name of the island to *Ebusus*. In 123 BCE the island was annexed to the Roman Republic, but maintained its cultural and political independence <sup>5</sup>; it was after the foundation of the Roman Empire in 27 BCE that the island lost its independence to become a federate city <sup>6</sup>. Later, around 902 CE the Umayyads of Cordoba occupied the island for about four centuries and Ibiza became *Yebisah* <sup>6</sup>, going through a process of islamization, as suggested by archaeological records <sup>7</sup>. In 1235, the invasion of the Catalans represented the last significant event in the colonization history of the island before it became the privileged destination of extensive immigration from the *hippy* community in the second half of the 20th century, paving the way for Ibiza to become the cosmopolitan island it is today. All of these events must have contributed in shaping the genetic background of Ibizans, but their impact on the modern islanders' genome remains unknown. A previous study based on classical genetic markers proposed a genetic continuity of the islanders with North African and East Mediterranean populations <sup>5</sup>. A Carthaginian maternal founder effect was also hypothesized based on analyses of the mitochondrial hypervariable region I (HVRI) <sup>8</sup>. However, no reduction of diversity was detected for Y-chromosome STRs, as compared with other Mediterranean groups <sup>9</sup>. In this study we aim to explore genomic diversity by conducting the first genome-wide analysis of the modern Ibizan population. We also aim to investigate whether Phoenician ancestry can be identified in modern Ibizans or whether this ancestry was lost due to demographic replacement and subsequent isolation.

## **MATERIALS AND METHODS**

**Dataset and genotypes.** Informed consent was given by 163 volunteers from three regions in Spain (Catalonia, Valencia, and the Balearic Islands; as a shorthand, we will refer to these as the *core regions* of our study); Internal Review Board approval for this work was granted by CEIC-PSMAR ref. 2016/6723/I. This research was conducted under the principles of the Helsinki declaration. The geographical origin of each subject was characterized and recorded as the average coordinates of the birthplaces of their four grandparents, which were distant at most 50 Km. For some analyses, a finer-grained geographical scale was used (by island in the Balearic Islands, by province in Valencia, and *regió*, namely the first-level administrative division used by the Catalan government, plus the city of Barcelona in Catalonia). DNA samples were collected from saliva and extracted as described in Solé-Morata et al. 2015 (ref. 10). Overall, four Axiom<sup>®</sup> Genome-Wide Human Origins Arrays (~629K SNPs)<sup>11</sup> were genotyped at the Centro Nacional de Genotipado - Universidade de Santiago de Compostela facility. Genotype calling was performed applying the Affymetrix Best Practices Workflow implemented in the software Axiom<sup>™</sup> Analysis Suite 2.0. Since the samples were distributed in different plates, two separate batches were run to avoid any batch effect during the genotype calling. A total of 41 samples failed the genotyping process, leading to 122 remaining individuals (Table S1). An Identity-by-descent (IBD) analysis was performed in order to unveil close relatedness between the donors. Two pairs of samples, both from the island of Ibiza, did not pass the test. Eventually, a total of 120 samples were retained (data available at [https://figshare.com/articles/People\\_from\\_Ibiza\\_dataset/7162025](https://figshare.com/articles/People_from_Ibiza_dataset/7162025)). Out of the 633,995 SNPs interrogated by the array, a different number of variants per batch passed the genotyping process (Figure S1). After the genotype calling, and before the quality

control filtering, the initial overlap between the four plates included 569,999 SNPs. The latter decreased to 519,297 variants when a further 69 Spanish samples from freely available data <sup>12</sup> were added. Thus, the final dataset increased to 189 samples, covering 13 out of the 17 regions in Spain (Figure S2). An ancient sample (MS10614) from the Phoenician necropolis of Cas Molí, on Ibiza, was included in the work. Genome-wide data for this sample were produced as described in Zalloua et al.<sup>13</sup>. Using 105 modern samples <sup>12</sup> from Lebanon, Syria, Israel, and Jordan, we reconstructed the ancient region of Canaan that we then merged together with our subjects from the core regions. Further 157 subjects <sup>14</sup> from Morocco, Algeria, and Tunisia were used for testing North African ancestry.

**Data Quality Control with PLINK.** X-chromosome and uniparental markers were excluded using the flag `--not-chr` implemented in PLINK<sup>15</sup> 1.90. Samples showing more than 10% of SNP genotypes missing were removed. We also rejected SNPs missing in more than 95% of the samples, and SNPs with extreme deviation from the Hardy-Weinberg equilibrium ( $p < 10^{-5}$ ), which left 519,249 SNPs; removal of SNPs displaying little variation in the sample set ( $MAF < 0.05$ ) resulted in 348,443 remaining SNPs. A set of independent SNPs was generated by pruning for linkage disequilibrium using the `--indep-pairwise` flag (the window size was set to 200 SNPs with a shifting of 25 SNPs, and  $r^2$  set to 0.5). We then performed an Identity-by-descent (IBD) analysis using the flag `--genome` in order to detect those samples with cryptic relatedness. A cutoff was applied up to the third degree of relationship ( $PI\_HAT > 0.125$ ), and for each related couple of samples the one showing the lowest SNP call rate was removed. Only the common SNPs to all the merged datasets were retained. After the quality control, polymorphism and LD filters in the global dataset of 189 samples, 142,520 common

SNPs passed and were used in further analyses. A different number of variants were achieved when considering the dataset with 120 samples from the three core regions (151,286 SNPs remaining after QC). In the analysis with the North African subjects, 55,969 final variants were used, while 47,057 SNPs were retained in the analysis that included the Levantine regions and the ancient sample.

**Principal Component Analysis (PCA).** The SmartPCA program (v. 13050) in EIGENSOFT<sup>16</sup> (v. 6.0.1) was used for computing the eigenvectors, and the results were plotted in R (v. 3.0.1).

**Fixation index ( $F_{ST}$ ).** The pairwise differentiation was computed between the different groups relying on the Weir and Cockerham's  $F_{ST}$  measure<sup>17</sup> implemented in the R package StAMPP<sup>18</sup>. Using the function `stampFst`, we bootstrapped across loci (1000 repetitions) in order to estimate the P-values associated to each  $F_{ST}$  measure. Also, a percentile method was implemented in order to compute the confidence intervals; a standard value of 95% was set to define the upper and the lower confidence intervals. Heatmaps were produced in RStudio<sup>19</sup> using R version 3.4.4 (ref. <sup>20</sup>). The Canary Islands sample was not included in this and all subsequent analyses given its extremely low sample size (N=2).

**Mantel test.** A geographic distance matrix was produced using The Geographic Distance Matrix Generator (v. 1.2.3, available from [http://biodiversityinformatics.amnh.org/open\\_source/gdmg](http://biodiversityinformatics.amnh.org/open_source/gdmg)) and, together with the  $F_{ST}$  matrix, was used as input for the Mantel test calculation using the `ade4` (ref. <sup>21</sup>) library in R. The plot was also generated with R using both `reshape`<sup>22</sup> and `ggplot2` (ref. <sup>23</sup>) libraries.

**ADMIXTURE.** Individual ancestries were computed with ADMIXTURE (v. 1.3.0)<sup>24</sup> and tested from K=2 up to K=8. A consensus for every K was calculated out of 10 independent runs using CLUMPP<sup>25</sup> (v. 1.1.2), while Distruct<sup>26</sup> (v. 1.1) was used for plotting the results for each one of the Ks. Furthermore, for K=2 up to K=5, each component was individually displayed along the geographic coordinates of the samples, resulting in the geographic distribution of the admixture percentages. These latter plots were obtained using the R function *tessplot* implemented in the *plot.admixture.r* script (freely available at <http://membres-timc.imag.fr/Olivier.Francois/plot.admixture.r>). Admixture f3 statistics were computed using the *threepop* function implemented in TreeMix<sup>27</sup>.

**Runs of homozygosity (ROH).** Pruned data underwent ROH analysis performed in PLINK. We scanned the genome using sliding windows containing at least 50 SNPs and a minimum length of 500 kb. At most, 1 heterozygous call and 5 missing calls per window were allowed. A threshold of 0.05 was set to define the proportion of homozygous windows overlapping the position of each SNP; at least one SNP per 50 kb was chosen as required minimum density value, and two consecutive SNPs were included in the same ROH when a maximum gap of 100 kb split them up. For each group, the average ROH number was calculated dividing the total number of runs (NROH) by the number of samples. The average total ROH length (namely, the average SROH per sample in each group) was calculated as the total ROH length (SROH) divided by the total number of individuals. The FROH (amount of autosomal genome covered by ROHs) for runs greater than 1 Mb was calculated according to the formula presented in McQuillan et al. 2008 (ref. <sup>28</sup>):

$$FROH_{>1Mb} = \frac{\sum L_{ROH}}{L_{auto}}$$



In this formula,  $\sum L_{ROH}$  is the  $S_{ROH>1Mb}$  calculated for each individual, and  $L_{auto}$  is a factor representing the length of the autosomal genome covered by SNPs on the array data (2,782.7 Mb).

**IBDseq/IBDNe.** In order to detect pairwise IBD blocks, we ran the program IBDseq<sup>29</sup> with default settings for the global dataset of 189 samples. We then extracted from the output the results for each group and used those ones as input for IBDNe<sup>30</sup> program. The latter allowed us to estimate the variation of  $N_e$  through the past few generations since we defined longer IBD segments by setting a cutoff at 7 cM (minibd=7). Apart from the minibd option that we customly set, we ran IBDNe with default settings, filtering the IBD segments as suggested by Browning & Browning<sup>30</sup>. Also, according to the software settings, 50 generations is the most reliable range we can take into account when working with SNP array data. We set a threshold of 100 IBD segments as the minimum sample size for this analysis, which was met only by the Ibizan and Catalan samples (440 and 223, respectively).

## RESULTS

### *Genetic differentiation in a Spanish context*

A first Principal Component Analysis (PCA) was performed considering the entire dataset of 189 Spanish samples (120 new plus 69 previously published). In the plot (Figure 1), the first PC clearly separates all of the Ibizan samples from the rest of the Spanish subjects, while the second PC separates Basque samples from the rest, within which no structure can be discerned. Because PC1 and PC2 explain approximately the same proportion of variance<sup>31</sup>, these results imply that the genetic differentiation of Ibiza from the rest of Spain is at least of the same order of magnitude as that of the

Basque Country. When we narrowed the analysis down to 120 samples in the three core regions (Catalonia, Valencia, and the Balearic Islands) we still detected two separate clusters defined along the first PC (Figure S3). Next, we quantified the genetic variation between groups through  $F_{ST}$ . Apart from Ibiza and the Basque Country,  $F_{ST}$  ranged from  $4 \times 10^{-5}$  to 0.002, while population pairs containing Ibiza had  $F_{ST}$  values ranging from 0.006 to 0.01 (Figure S4a and Table S2). Within the core regions,  $F_{ST}$  was even more homogeneous, ranging from  $3 \times 10^{-6}$  to 0.002, while Ibiza showed higher values of differentiation (0.006 - 0.007, Figure S4b). A Mantel test between the  $F_{ST}$  values and the geographic distances showed a positive but not significant correlation for both the global dataset ( $R^2=0.01$ ,  $P=0.4$ ), and within the core regions ( $R^2=0.164$ ,  $P=0.2$ ). Interestingly, when Ibiza was excluded from the test, the positive correlation became significant in the core regions, with  $R^2=0.353$  and  $P=0.009$  (Figure S5a). This is likely a reflection of the genetic outlier position of Ibiza, since its differentiation is larger than expected based on its geographical position, to the point that its presence in the analysis dramatically decreases the correlation that exists between genetic and geographic distances when Ibiza is excluded. On the other hand, the global dataset maintained a p-value  $> 0.05$  (Figure S5b), which can be attributed to the presence of another genetic outlier, namely the Basques.

In order to determine the global ancestry components of our samples, an ADMIXTURE analysis was performed, in which we found that  $K=2$  presented the lowest cross-validation error. The two major components detected (Figure 2) mirror the PCA results: one component prevails in Ibiza, the other in the Basque Country, and all other individuals contain on average 31.78% of the first vs. 68.22% of the second. Notably, moving through the different numbers of theoretical ancestral populations tested, we see

a consistent pattern in the fraction of the component coming from Ibiza (with an average of 87.45% along the eight Ks tested), while the other source starts to break up at K=3, as it is shown in the Admixture-map plots (K=2 up to K=5 shown in Figure S6).

#### *Comparison with an ancient Phoenician sample*

A sample (MS10614) from the Phoenician Cas Molí site in Ibiza<sup>13</sup> was sequenced at a 0.47X coverage, and was projected into a PCA space with two modern groups of samples, from Iberia (including contemporary Ibiza) and the Levant (Figure S7a). The ancient Phoenician sample clusters with the modern Levantines, rather than with the cluster represented by the core regions. Furthermore, the admixture proportions seem to be consistent with this result (Figure S7b). In the figure, the average percentage for each component is displayed for each group; it is clear that the ancient sample is closely related to the groups from Levant. And, conversely, these results seem to be incompatible with direct continuity between the current population of Ibiza and the ancient Phoenicians who settled the island.

#### *Admixture into Ibiza*

When we tested Ibiza for both East Mediterranean and North African ancestries, we did not find any major signal of differentiation from the rest of the Spanish groups. In the admixture analysis with the 105 Levantine subjects, K=2 turned out to be the most suitable model according to the cross-validation error analysis; the two components we detected can be clearly attributed to the Iberian Peninsula and the Levant, respectively (Figure S7b). Concerning the North African influence, a principal component analysis clearly shows a separation of the three core regions from the North African samples

(Figure S8a). This split is also corroborated by the admixture results (Figure S8b); even in this case,  $K=2$  describes the structure of the groups highlighting the presence of two distinct components. As in the PCA, Ibiza conforms to the rest of the Spanish groups. The absence of any signal of admixture in the Ibizan population was also supported by  $f_3$  statistics; in figure S9 we show the results for all the tested trios, including Northern Africans, Middle Easterners, and also the ancient sample from Ibiza.

#### *Endogamy and isolation in Ibiza*

Endogamy in small population isolates and inbreeding leave their mark in the genome as long runs of homozygosity (ROHs). When we separated the ROHs according to different length categories as in Kirin et al.<sup>32</sup>, it was clear that Ibiza stood out for both average ROH number, and average total ROH length for the interval 1-2 Mb (Figure 3a and 3b), and was second only to Cantabria (with a huge dispersion) in the interval 2-4 Mb. Possibly, the scarce number of samples for some groups in this category of ROH can explain this contingency; indeed, the proportion of individuals in Ibiza carrying ROHs in this size range is 76%, while it is at most 50% in any other population (Figure 3c). The observed ROH configuration in Ibiza seems to fit the very recent history of modern islanders; short ROHs are common in populations that have experienced a bottleneck, reflecting also a more recent parental relatedness<sup>33</sup>. In our analysis, Ibiza shows the highest levels of the total length of ROHs (SROH, Figure S10a) with the subjects showing a homogeneous distribution through the different cumulative lengths, while different groups show outliers with values well further away from the mean. As for the total number of ROHs (NROH), Ibiza is second only to the Basque Country

(Figure S10b), as we can also appreciate in Figure S11a in which the Ibizan samples are more widely spread along the two axes than any other group. On the other hand, the majority of the other groups mass in an area of the plot represented by fewer and shorter ROHs. Interestingly, samples from the Basque Country show an intermediate distribution along the axes, with some samples from the Catalan group spread through higher values, mirroring the same pattern observed for the outliers seen in the whisker plots (for a clearer visualization, refer to Figure S11b)). Furthermore, in Figure S10c it is possible to appreciate how the subjects from Ibiza stand out from the others for the proportion of the autosomal genome covered by runs of homozygosity (FROH). On average, each subject from Ibiza displays 0.81% of the genome as autozygous, the highest value retrieved among all the groups (Table S3).

#### *Demographic history*

By using IBDseq and IBDNe, we calculated a series of effective population size ( $N_e$ ) values through the last 50 generations from the present. Out of all the pairwise IBD segments retrieved with IBDseq, only a limited number passed the filter for minimum allowed length set in IBDNe ( $\text{minibd}=7$  cM), with Ibiza showing the highest fraction of passing IBD segments (20.27%).  $N_e$  through time is plotted in Figure 4; for Ibiza, the general trend in the last 1500 years is clearly a reduction of the effective population size, up to a minimum  $N_e$  found 10-15 generations in the past, followed by an exponential growth in more recent times. When comparing this pattern with that found in the mainland Catalan population, we detected a general similar trend but higher  $N_e$  values, and a faster increase of the slope after the observed minimum. We used the IBDseq output in order to quantify the amount of IBD chunks that each sample from

Ibiza receives on average from the other groups (Figure S12a). In this analysis we subsampled both Catalan and Valencian groups in order to avoid any sample size bias. We found that each subject from Ibiza receives the highest amount of IBD chunks from its own group. Furthermore, we calculated the total length (bp) of shared IBD chunks, and represented how much on average, each Ibizan sample receives from every other population (Figure S12b); the results highlight that Ibizans receive on average the longest IBD chunks from the Ibizan group itself.

## **DISCUSSION**

Our study suggests that the population of Ibiza is genetically distinct from the rest of Spain, a similar phenomenon to what has been observed with the Basques who have been previously described as a genetic outlier<sup>34,35,36</sup>. The different approaches we applied in this study support the fact that the island is a genetic isolate. The results we achieved actually reflect the same pattern observed for other known isolated populations. For example, the principal component analysis clearly points out that Ibiza forms its own cluster, recalling the way Sardinia splits from the rest of the Italian Peninsula<sup>37</sup>. This separation is also supported by the genetic variation analysis between groups. Indeed, even if a general homogeneity is detected, Ibiza shows higher values of differentiation. Following the parallelism with Sardinia, the Italian island diverges from other European groups for  $F_{ST}$  values<sup>34</sup> that, interestingly, mirror the ones that we observed when comparing Ibiza to the different regions of Spain. Furthermore, it is of particular interest to notice that the value detected for the genetic distance of Ibiza from the Basque Country recalls the one that Sardinia shares with that particular community ( $F_{ST} = 0.011$ )<sup>34</sup>. The latter observation was also supported by a further  $F_{ST}$  analysis (data not shown) including Sardinian samples genotyped with the Human Origins Array from

Lazaridis et al.<sup>12</sup>; the results were similar to those observed in previous studies<sup>34</sup>. Even when we look at the admixture results, the pattern displayed at  $K=2$  underscores the presence of a strong signature component that Ibiza preserves through all the  $K$  values we tested, without showing any significant break. This is different from the behavior displayed by any of the other groups. The nature of this signature component as a signal of isolation seems to be a very common pattern observed in other genetic isolates, such as the Kalash<sup>38</sup>, and the Sardinians<sup>39</sup>. Furthermore, the configuration observed for both the Ibizan and the Basque components from  $K=3$  on, could be the result of a genetic drift-driven redistribution of the allele frequencies, which seems to be common in populations that experienced a rapid size reduction<sup>40</sup>. Since genetic differentiation can be the consequence of very different events, we tested various hypotheses for the reasons behind the differentiation of modern Ibizans.

#### *Geographical isolation.*

The hypothesis that modern Ibizans differ from continental Spanish populations just because they live far from them has been taken into account. A Mantel test highlighted a significant correlation between the geographical distances and the genetic ones, only when Ibiza was removed from the analysis ( $R^2=0.353$ ,  $P=0.009$ ). This can be taken as an indication of the genetic outlier position of Ibiza given that its presence breaks the correlation between genetic and geographic distances; note that this analysis also contains the other Balearic Islands (Mallorca and Minorca), the latter of which is similar in area and population size to Ibiza, and is further away from the mainland. Thus, the insularity and geographical position of Ibiza do not seem to be major contributors to its genetic distinctiveness.

*Are Ibizaans descended from the Phoenician settlers in the island?*

We tested whether Phoenicians, represented by an ancient DNA, low-coverage genome, had contributed to the genetic make-up of present-day islanders, and tested signals of admixture with modern Eastern Mediterranean and North African populations. All the archeological evidence to date suggest that it was the Phoenicians who first settled Ibiza, giving it a central role in the Western Mediterranean trade network, and making it “one of the true cities of the Western European antiquity”<sup>6</sup>. With five centuries of colonization, it has long been suggested that the legacy of the ancient Phoenicians might be contained in the genomes of current Ibizaans<sup>5,8,9,41</sup>. We have had the opportunity to properly explore this possibility because of the availability of an ancient Phoenician genome from Ibiza. As pointed out by the results, it is evident that there is no genetic continuity between the modern inhabitants of the island and this ancient subject. Furthermore, our results exclude any major Middle Eastern specific component in modern Ibizaans, or any other Spanish population. Nevertheless, it is surprising how the modern Levantines still share a genetic continuity with the ancient Ibizan (MS10614), pointing to the possibility that modern samples from those regions might be considered as proxies of the founding civilization. Also, a greater similarity to the rest of the Iberians rather than to Levantines in this analysis suggests that the modern Ibizan differentiation is not caused by any other Ibizan-specific external admixture contribution. In support of this observation, our results also exclude any major North African component in modern Ibizaans, as well in the rest of the Spanish subjects. Also, testing more specifically for admixture, the  $f_3$  statistics gave no signal of admixture when we tested the Ibizan group against Northern Africans, Middle Easterners, and the



ancient sample. According to historical and archaeological records, apart from the European ones, Ibiza owes its ancestry to East Mediterranean and North African populations<sup>5</sup>; our results support the idea that, even if any external influence occurred, it has not persisted to the current Ibizan population.

#### *Ibiza: a demographic history*

The exclusion of any major genetic Phoenician influence in the modern Ibizan genome led us to explore the demographic history of the islanders considering the effects of consanguinity, given that it has long been a common practice on the island (marriages between relatives up to the 4<sup>th</sup> degree were very frequent, and there are records of unions between second-degree cousins dated back to the late 1960s<sup>42</sup>). This part involved the analysis of ROHs; runs of uninterrupted stretches of the genome, inherited identically from both parents, which can be used to unveil both recent and ancient relatedness between subjects. Normally, longer ROHs are associated with recent inbreeding because of the lack of recombination in the short term, while shorter ROHs are symptomatic of a much older story<sup>32</sup>. The outcome of our analyses highlights the presence of more, and longer ROHs in Ibiza than in any other group, a demographic scenario which is similar to the one that was previously associated to inbred populations<sup>28</sup>, or in general to small groups, in which the number and the length of ROHs are usually greater than those observed in larger populations<sup>33</sup>. Moreover, the high number of short ROHs some of the samples show is usually symptomatic of a population bottleneck event<sup>33</sup>. However, the scenario depicted for the majority of the other Spanish groups is typical of large, non-isolated populations, characterized by few and shorter ROHs<sup>28</sup>. Furthermore, the samples from Catalonia, Valencia, and the remaining Balearic Islands (Majorca and Menorca) present some individuals with larger values for

both NROH and SROH. These outliers, with atypically long haplotypes, may be caused by a low recombination rate in some genomic regions that has led to the appearance of misleadingly long ROHs in outbred populations<sup>32,43,44</sup>. However, while the possibility that the haplotypes carried by these outliers could have arisen from recent parental relatedness cannot be totally excluded, the fact that the general trend for these groups is the one usually observed in non-isolated populations seems to be the most reasonable explanation. The Basque Country samples represent a different situation, with a combination of both shorter and longer ROHs that suggests a more complex history. On average, the subjects from Ibiza show the highest number of ROHs for the length category 1-2 Mb, and probably for the 2-4 Mb as well (if we consider that other populations have a rather low sample size of individuals carrying ROHs in this range). Typically, ROHs with lengths 1-2 Mb fall in a category associated to a background of parental relatedness due to limited population size<sup>45</sup>. Even though consanguineous unions are not frequently observed in European populations<sup>32,46</sup>, in the Mediterranean area the highest percentage is recorded in Spain (3.5%)<sup>46</sup>. Furthermore, ROHs with lengths 2-4 Mb have been linked to ancient parental relatedness resulting from genetic drift and frequent consanguineous practices<sup>47</sup>. One of the consequences of inbreeding is the inheritance of alleles that are identical by descent, resulting in the increase of the subjects' overall autozygosity; the higher this value is, the greater the chance that a population has experienced some sort of inbreeding event. Our analysis on the degree of individual autozygosity (FROH) clearly pointed out that Ibizans possess the highest autosomal coverage in ROHs. Overall, our results are consistent with the history of consanguineous unions on the island. Current Ibizans are descendants of the Catalan conquerors of the 13th century, after which the island became a socio-geographic

isolate, promoting, especially in rural areas, the spread of the relatively stable and impervious practice of marriages between relatives<sup>48</sup>. One of the consequences of this habit is undoubtedly the reduction of genetic diversity and the consequent increased isonymy (decrease in the recorded variability of surname combinations<sup>42</sup>). Following the ROH results, the analysis of IBD segments detected a recent reduction in the population size due to a population bottleneck. The subjects from Ibiza have shown the highest number of IBD segments, with significant statistical support in terms of LOD scores retrieved with IBDseq. Through IBDNe, we observed how the effective population size of the island changed in the last 50 generations, with a significant collapse within the last 10-15 generations. Even if IBDNe for small datasets is prone to loss of precision in the most recent generations<sup>30</sup>, historical records seem to fit well with the pattern observed; the population decline we detected may correspond to a period from the second half of the 16th century to the beginning of the 18th century; during this period, the island suffered from famine after the Franco-Ottoman attack in 1536, and was harshly impacted by the global bubonic plague epidemic in 1652 (ref.<sup>49</sup>). Our results suggest a recovery in the population size beginning around the early 18th century, which is consistent with the historical record when Ibiza swore allegiance to the crown of Philip V of Spain in 1715. Compared to Ibiza, Catalonia shows an even more rapid increase in population size, after a similar period of reduction; this too is consistent with the historical records, since in 1651, in addition to being under siege by the Castilian troops in the Reapers' War, a third of the population of Barcelona was devastated by the bubonic plague<sup>49</sup>. Moreover, in figure S12a we highlight how on average each subject from Ibiza receives more IBD chunks from itself and then from the other groups, while the longest IBD chunks clearly come from the Ibizan group (figure

S12b). These results support the increased genetic drift in Ibizans and their genetic isolation from other Spanish groups. However, detecting any founder effect from the Catalan source through the analysis of the IBD sharing seems to be unlikely at a such small-scale. This is possibly due to the highly drifted Ibizan sample, and the high homogeneity among the mainland populations. In conclusion, our study suggests that the elevated genetic differentiation Ibizans have from other Spanish groups does not result from genetic continuity with Phoenician founders; also, the presence of more and longer ROH segments supports the hypothesis of increased parental relatedness due to the well recorded practice of consanguineous unions on the island, while the IBD chunks distribution supports the genetic drift and also suggests that a population collapse, 10-15 generations ago, affected the island, likely corresponding to the bubonic plague occurred in 1652.

## **ACKNOWLEDGEMENTS**

We thank Maria Ferrer (IBE, CSIC-UPF) for her invaluable help in finding the Ibiza sample donors, and Inés Quintela (CEGEN - USC) for her assistance in genotyping the samples. Of course, this work would not have been possible without the kind collaboration of all the sample donors. Funding was provided by the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (FEDER) (grant CGL2016-75389-P), Agència de Gestió d'Ajuts Universitaris i de la Recerca (Generalitat de Catalunya) grant 2014 SGR 866, and “Unidad de Excelencia María de Maeztu”, funded

by the MINECO (ref: MDM-2014-0370). SAB was supported by the Agencia Estatal de Investigación FPI grant BES-2014-069224.

### **Conflict of interest statement**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

### **REFERENCES**

- 1 Somers M, Olde Loohuis LM, Aukes MF *et al.* A genetic population isolate in the Netherlands showing extensive haplotype sharing and long regions of homozygosity. *Genes (Basel)* 2017; **8**: 1–14.
- 2 Braudel F. *Une Leçon d'histoire*. 1985.
- 3 Institut d'Estadística de les Illes Balears (IBESTAT). IBESTAT (2017). Cifras de població. .
- 4 Ribas BC. Eivissa-Prehistòria. Enciclopèdia d'Eivissa i Formentera. 2006.
- 5 Picornell A, Ana M, Castro JA, Ramon MM, Arya R, Crawford MH. Genetic Variation in the Population of Ibiza (Spain): Genetic Structure, Geography, and Language. *Hum Biol* 1996.
- 6 Armstrong S. *The White Island*. 2004.
- 7 Gurrea Barricarte R, Martín Parrilla Á. Eivissa Història Època califal. Enciclopèdia d'Eivissa i Formentera. 2015.
- 8 Picornell A, Gómez-Barbeito L, Tomàs C, Castro JA, Ramon MM. Mitochondrial DNA HVRI variation in Balearic populations. *Am J Phys Anthropol* 2005; **128**: 119–130.

- 9 Tomàs C, Jiménez G, Picornell A, Castro JA, Ramon MM. Differential maternal and paternal contributions to the genetic pool of Ibiza Island, Balearic Archipelago. *Am J Phys Anthropol* 2006; **129**: 268–278.
- 10 Solé-Morata N, Bertranpetit J, Comas D, Calafell F. Y-chromosome diversity in Catalan suRNAme samples: Insights into suRNAme origin and frequency. *Eur J Hum Genet* 2015; **23**: 1549–1557.
- 11 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y. Ancient Admixture in Human History. *Genetics* 2012; **192**: 1065–1093.
- 12 Lazaridis I, Patterson N, Mittnik A *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 2014; **513**: 409–413.
- 13 Zalloua P, Collins CJ, Gosling A *et al.* Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Sci Rep* 2018; **8**: 17567.
- 14 Arauna LR, Mendoza-Revilla J, Mas-Sandoval A *et al.* Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. *Mol Biol Evol* 2017; **34**: 318–329.
- 15 Purcell S, Neale B, Todd-Brown K *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 16 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: 2074–2093.
- 17 B. S. Weir and C. Clark Cockerham. Estimating F-Statistics for the Analysis of Population Structure. *Evol Vol* 38 1984; **38**: 1358–1370.
- 18 Pembleton LW, Cogan NOI, Forster JW. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol*

- Ecol Resour* 2013; **13**: 946–952.
- 19 RStudio Team. RStudio Team (2015). RStudio: Integrated Development for R. 2015.<http://www.rstudio.com/>.
- 20 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2018.
- 21 Dray S, Dufour A-B. The **ade4** Package: Implementing the Duality Diagram for Ecologists. *J Stat Softw* 2007; **22**. doi:10.18637/jss.v022.i04.
- 22 Wickham H. Reshaping Data with the `\pkg{reshape}` Package. *J Stat Softw* 2007; **21**: 1–20.
- 23 Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Media* 2009; **35**: 211.
- 24 Alexander DH, Novembre J. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res* 2009; **19**: 1655–1664.
- 25 Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 2007; **23**: 1801–1806.
- 26 Rosenberg NA. DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 2004; **4**: 137–138.
- 27 Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* 2012; **8**. doi:10.1371/journal.pgen.1002967.
- 28 McQuillan R, Leutenegger AL, Abdel-Rahman R *et al*. Runs of Homozygosity in European Populations. *Am J Hum Genet* 2008; **83**: 359–372.
- 29 Browning BL, Browning SR. Detecting identity by descent and estimating genotype error rates in sequence data. *Am J Hum Genet* 2013; **93**: 840–851.

- 30 Browning SR, Browning BL. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *Am J Hum Genet* 2015; **97**: 404–418.
- 31 McVean G. A genealogical interpretation of principal components analysis. *PLoS Genet* 2009; **5**. doi:10.1371/journal.pgen.1000686.
- 32 Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 2010; **5**: e13996.
- 33 Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: Windows into population history and trait architecture. *Nat Rev Genet* 2018; **19**: 220–234.
- 34 Rodríguez-Ezpeleta N, Álvarez-Busto J, Imaz L *et al*. High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. *Hum Genet* 2010; **128**: 113–117.
- 35 Comas D, Mateu E, Calafell F *et al*. HLA class I and class II DNA typing and the origin of Basques. *Tissue Antigens* 1998; **51**: 30–40.
- 36 Calafell F, Bertranpetit J. Principal component analysis of gene frequencies and the origin of Basques. *Am J Phys Anthr* 1994; **93**: 201–215.
- 37 Fiorito G, Di Gaetano C, Guarrera S *et al*. The Italian genome reflects the history of Europe and the Mediterranean basin. *Eur J Hum Genet* 2016; **24**: 1056–1062.
- 38 Ayub Q, Mezzavilla M, Pagani L *et al*. The kalash genetic isolate: Ancient divergence, drift, and selection. *Am J Hum Genet* 2015; **96**: 775–783.
- 39 Di Gaetano C, Fiorito G, Ortu MF *et al*. Sardinians genetic background explained



- by runs of homozygosity and genomic regions under positive selection. *PLoS One* 2014; **9**: 1–8.
- 40 Listman JB, Malison RT, Sughondhabirom A *et al.* Demographic changes and marker properties affect detection of human population differentiation. *BMC Genet* 2007; **8**: 21.
- 41 Adams SM, Bosch E, Balaesque PL *et al.* The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 2008; **83**: 725–736.
- 42 Claudio Alarco von Perfall. Sobre los matrimonios consanguíneos en Ibiza. *Eivissa, num8* 1976; **16**: 328–331.
- 43 Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* 2006; **15**: 789–795.
- 44 Christofidou P, Nelson CP, Nikpay M *et al.* Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages. *Am J Hum Genet* 2015; **97**: 228–237.
- 45 Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 2012; **91**: 275–292.
- 46 Ceballos FC. La genética de los matrimonios consanguíneos. *Dendra Médica Rev Humanidades* 2011; **10**: 160–176.
- 47 Jalkh N, Sahbatou M, Chouery E, Megarbane A, Leutenegger AL, Serre JL. Genome-wide inbreeding estimation within Lebanese communities using SNP arrays. *Eur J Hum Genet* 2015; **23**: 1364–1369.
- 48 Rozenberg D. Endogàmia i problemes d'identificació a Eivissa: l'exemple de

Sant Miquel. *Eivissa, num12* 1981; : 515–518.

49 Kohn GC. *Encyclopedia of Plague and Pestilence: From Ancient Times to the Present*. 1995.

**Titles and legends to figures.**

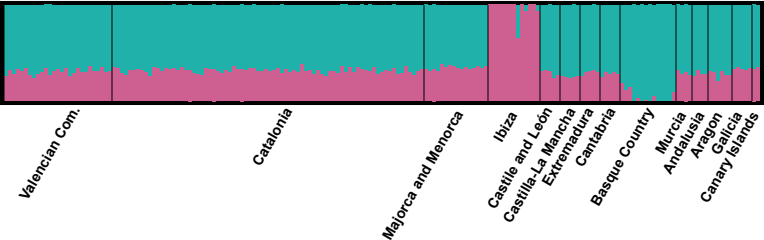
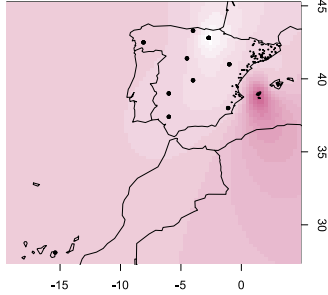
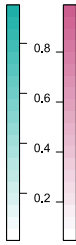
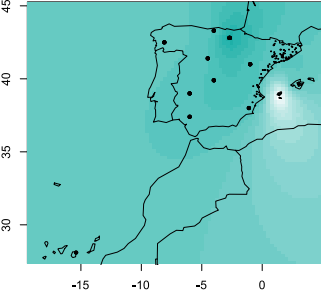
**Figure 1** Principal Component Analysis. Samples grouped according to the autonomous community they belong to. The geographic distribution for the different groups is displayed in the side map

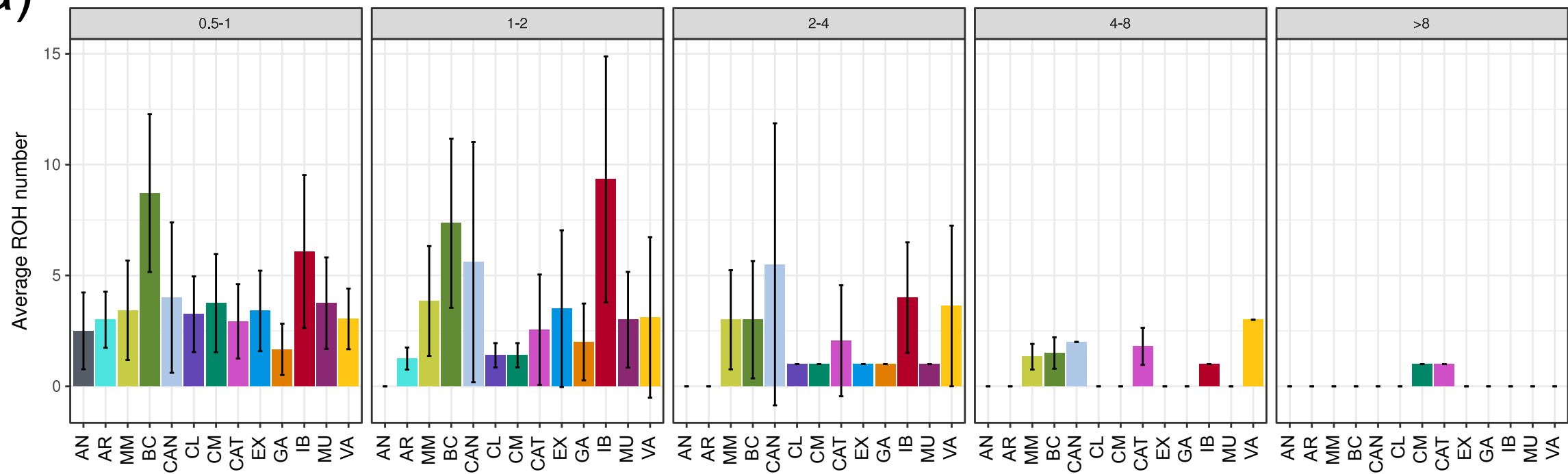
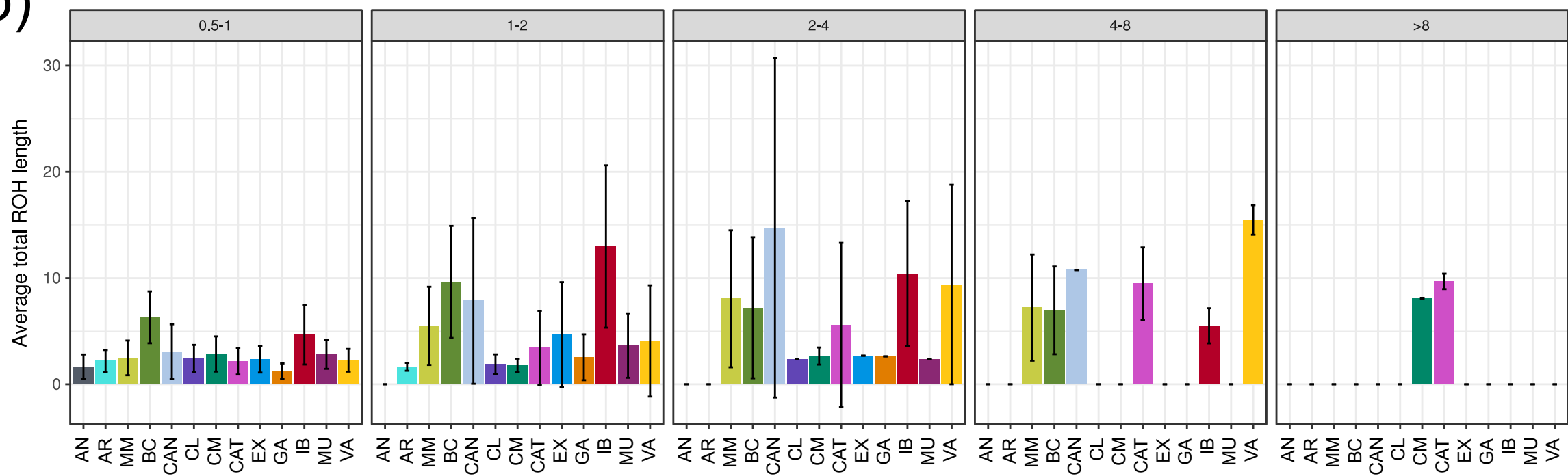
**Figure 2** ADMIXTURE results at K=2 for 13 autonomous communities (Ibiza is separated from the Balearic Islands since it is the target of this study). The maps above the ADMIXTURE bar plot represent the spatial distribution of the admixture coefficients. Side bars show the percentage distribution for each one of the components

**Figure 3** ROH distribution in five length categories according to (a) average ROH number, (b) average total ROH length, and (c) proportion of subjects having ROHs falling in each category. In (a) and (b), values have been averaged according to the number of the subjects showing that particular category of ROH in each population. Acronyms: AN, Andalusia; AR, Aragon; MM, Majorca and Menorca; BC, Basque Country; CAN, Cantabria; CL, Castile and León; CM, Castilla-La Mancha; CAT, Catalonia; EX, Extremadura; GA, Galicia; IB, Ibiza; MU, Murcia; VA, Valencia.

**Figure 4** Variation of the recent effective population size ( $N_e$ ) in the last 50 generations ( $g=30$  years) for Ibiza and Catalonia. Shaded area represents the upper and lower 95% confidence intervals. Vertical black line refers to the year 1652, known as notorious *annus horribilis* because of the last big bubonic plague. Even if the two trends are very similar, it is evident a clear difference in the  $N_e$  values.





**(a)****(b)****(c)**