

High Differentiation among Eight Villages in a Secluded Area of Sardinia Revealed by Genome-Wide High Density SNPs Analysis

Giorgio Pistis^{1,9}, Ignazio Piras^{2,9}, Nicola Pirastu², Ivana Persico¹, Alessandro Sassu², Andrea Picciau², Dionigio Prodi², Cristina Fraumene², Evelina Mocci², Maria Teresa Manias², Rossano Atzeni², Massimiliano Cosso², Mario Pirastu^{1,2}, Andrea Angius^{1,2,9*}

1 Istituto di Genetica delle Popolazioni, CNR, Alghero, Sassari, Italy, **2** Sardinia Life Sciences, Pula, Cagliari, Italy

Abstract

To better design association studies for complex traits in isolated populations it's important to understand how history and isolation moulded the genetic features of different communities. Population isolates should not "a priori" be considered homogeneous, even if the communities are not distant and part of a small region. We studied a particular area of Sardinia called Ogliastra, characterized by the presence of several distinct villages that display different history, immigration events and population size. Cultural and geographic isolation characterized the history of these communities. We determined LD parameters in 8 villages and defined population structure through high density SNPs (about 360 K) on 360 unrelated people (45 selected samples from each village). These isolates showed differences in LD values and LD map length. Five of these villages show high LD values probably due to their reduced population size and extreme isolation. High genetic differentiation among villages was detected. Moreover population structure analysis revealed a high correlation between genetic and geographic distances. Our study indicates that history, geography and biodemography have influenced the genetic features of Ogliastra communities producing differences in LD and population structure. All these data demonstrate that we can consider each village an isolate with specific characteristics. We suggest that, in order to optimize the study design of complex traits, a thorough characterization of genetic features is useful to identify the presence of sub-populations and stratification within genetic isolates.

Citation: Pistis G, Piras I, Pirastu N, Persico I, Sassu A, et al. (2009) High Differentiation among Eight Villages in a Secluded Area of Sardinia Revealed by Genome-Wide High Density SNPs Analysis. PLoS ONE 4(2): e4654. doi:10.1371/journal.pone.0004654

Editor: Mark A. Batzer, Louisiana State University, United States of America

Received: October 21, 2008; **Accepted:** January 29, 2009; **Published:** February 27, 2009

Copyright: © 2009 Pistis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Italian Ministry of Education, University and Research (MIUR) n°: 5571/DSPAR/2002 and (FIRB) D.M. n.718/Ric/2005.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: angius@igm.ss.cnr.it

⁹ These authors contributed equally to this work.

Introduction

Recent advances in array technologies have open up the possibility of economic and rapid genotyping of entire cohort of population samples. The availability of millions of single nucleotide polymorphisms (SNPs) provides a highly dense map across the human genome, which can achieve adequate power for investigating genome variants associated to multifactorial diseases [1].

Knowledge of the population structure and LD pattern is essential for study design in order to choose the appropriate approach for gene identification [2,3].

The identification of genetic variants underlying common human diseases are deeply affected by different population structures even if the research is very carefully designed.

In whole genome association studies, population stratification influences results when allele frequencies differ among subpopulations that are not represented equally among cases and controls. Undetected population structure can mimic the signal of association and can lead to false positives or to missed real effect [4–6]. Several studies demonstrate that sampling strategies need to take into account substructures even in relatively homogeneous

genetic isolates and this is even more relevant in inbred populations [7].

Population substructures have been detected in large population groups [8–9], but also in relatively homogeneous geographical and cultural genetic isolates as Iceland, Finland and Jewish Ashkenazi, that cannot be considered to be a single, randomly interbreeding population [7,10–12].

LD genome mapping is an important parameter for the design of association mapping studies [13,14]. It has been suggested that population isolates, particularly those founded recently, have longer stretches of LD than outbred populations [15]. The interest in utilizing population isolates for LD mapping studies has grown considerably, because increased LD values found in such populations, reduce the number of markers needed decreasing costs and improving statistical analysis [15]. Although isolates show a reduced allelic diversity [4], the presence of long-range LD regions could be useful to identify rare genetic diseases [16] more frequent in these populations.

Not all isolates are equal: marker informativity and extent of LD could vary substantially in different isolated populations [3,17,18]. Sardinian population could be considered a genetic homogeneous

isolate if one focuses on the founder effect causing monogenic diseases such as β -thalassemia, Wilson, etc. [19,20]. However, studies of the population genetic structure in different sub-regions of the island identified micro differentiation [21–23]. The comparison of small areas or isolated villages always revealed differentiation, while different results were obtained comparing larger areas [24–29].

The central-eastern area of Sardinia called Ogliastra, is characterized by the presence of several distinct villages with different history, immigration events and number of inhabitants [21,30]. Centenarian cultural isolation and conservatism characterized the history of these communities [31]. The relevant features of villages in this geographic area are similar environmental conditions, high endogamy, low immigration and remote origin [28,32]. This region also appears genetically differentiated from neighboring areas [21].

We focused our analysis on eight Ogliastra isolated villages using a wide set of about 360000 informative SNPs to determine the extend of genetic variation in this population. Other papers compared genetic features in isolated populations but they either used a limited number of SNPs or single chromosomes or have referred only to partial regions [1,15,33]. On the contrary, we wanted to examine population structure in all eight villages using an extensive subset of SNPs covering the entire genome.

Our aim was to understand how genetic features can be influenced by history and by geographical and cultural isolation in different communities in order to better determine the feasibility and design of genetic analysis for complex traits.

Materials and Methods

Sample characteristics

Historic and demographic data demonstrate that Sardinia experienced waves of successive invading populations that pushed the original inhabitants into the most remote and inaccessible areas. The central-eastern Sardinia region, identified as Ogliastra, is clearly one of these refuge areas, geographically and socially secluded for thousands of years from other Sardinia regions due to mountains and deep river valleys. Only after the Second World War, emigration and

economic opportunities have changed the population structure of some of these mountain villages leaving in place mostly the older generations of people who still reflect the high level of consanguinity of past inbreeding and who still follow the traditional life styles. To conduct our research we recruited about 9000 voluntary subjects in 8 selected villages: Talana, Urzulei, Baunei, Triei, Seui, Seulo, Ussassai and Loceri (figure 1). All are in the Ogliastra region except for Seulo (which is in Barbagia, another refuge area of Sardinia) but only 5 kilometers from Seui. Moreover, Seulo was in the past included in the Bishopric of Ogliastra and all of its parish registers (*Quinque Libri*) are also conserved in the same archive.

The 8 selected communities share similar demographic features: limited number of founders, high endogamy and consanguinity. We reconstructed their demographic dynamics over the centuries in order to assess their degree of isolation and inbreeding. Six small villages showed similar population growth trend from 1688 to 2001. Baunei and Seui differ because they were two large villages: Seui was one of the most important centers from the 17th century to 1931 because of anthracite and copper mines but today, because of emigration toward Cagliari, is comparable in size to the other small villages (figure 2).

We calculated the level of endogamy characterizing these populations from 1676 to 1975 (figure 3). Endogamy was calculated as the percentage of marriages' number among people from the same village divided by the marriages celebrated in the same 25 years interval [34].

The endogamy reaches a very high degree in each village from 1676 to 1950 with an average of 80.75% [min 67.18% (Loceri); max 92.72% (Baunei)], starting a slow progressive decrease from 1900 to 1950. From 1950, we observe a severe endogamy decline in all villages (average 57.03%) except for Talana (81.28%).

We reconstructed for each village extensive genealogies based on a relational database that includes all the inhabitants personal data starting from the 17th century. Based on kinship coefficients [35], we selected 45 unrelated individuals from each of the 8 villages forming a sample of 360 individuals. All selected individuals descent from families that were born and lived in the same village for at least four generations. Moreover, as comparison, we also subsequently carried out the same type of

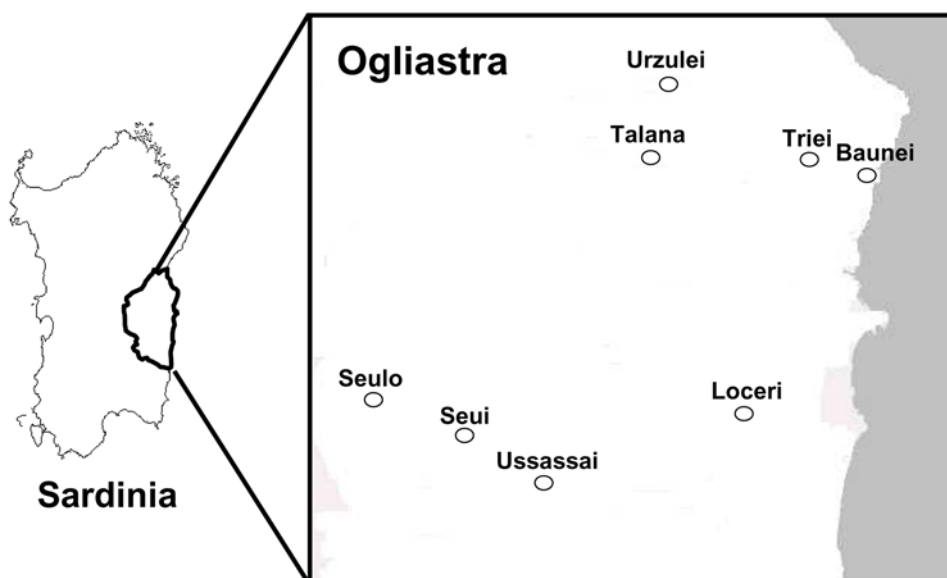


Figure 1. On the left, geographic localization of Ogliastra region in Sardinia. On the right, detailed location of the 8 villages analyzed.
doi:10.1371/journal.pone.0004654.g001

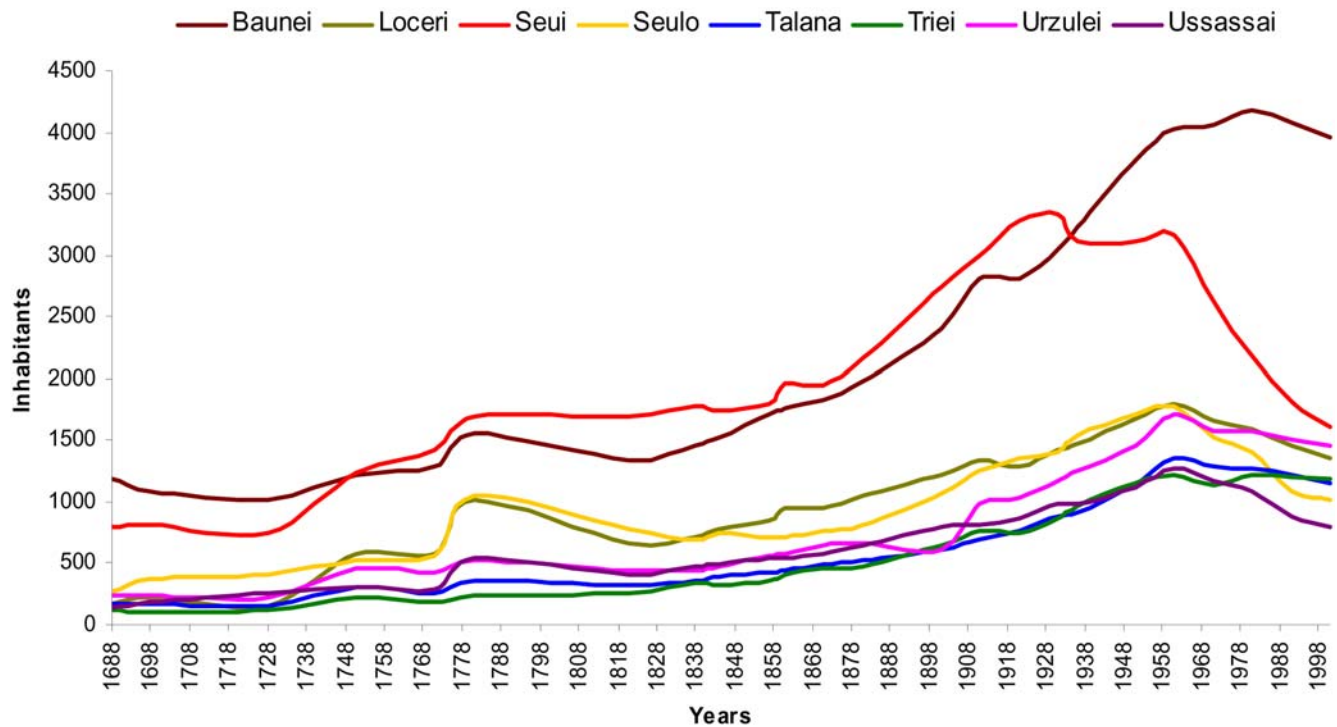


Figure 2. Population growth in the 8 villages analyzed starting from the 17th century to the year 2001.
doi:10.1371/journal.pone.0004654.g002

analysis (LDU, D' and r² estimates) on 43 individuals originating from other geographical areas of Sardinia.

People living in the village were invited to take part in the study by invitations sent by mail to every family and through public announcements. All the people participating in the study were healthy individuals and volunteers. The personal data were

encrypted and kept separated from genetic and data. The scientific content of the project was extensively explained to all individuals participating in the study that signed informed consent forms in accordance with the Helsinki Declaration. The study was submitted and approved by the Italian Ministry of University and Research (MIUR) following the current Italian legislation.

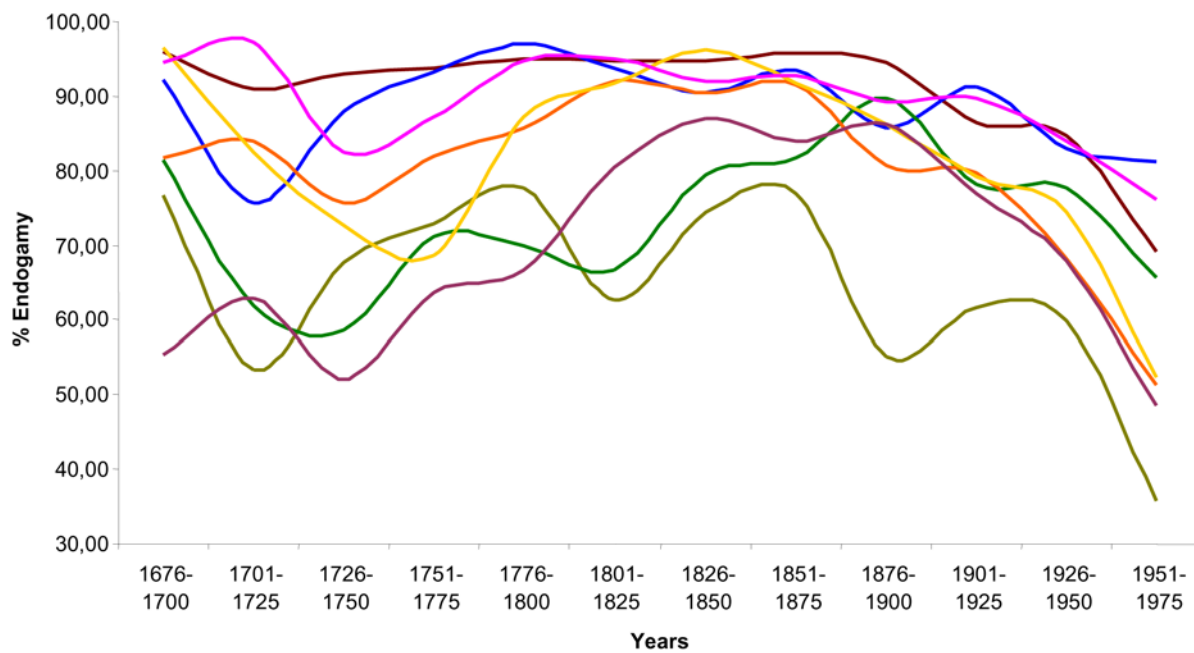


Figure 3. Endogamy (values in %) in the Ogliastra villages from 1676 to 1975.
doi:10.1371/journal.pone.0004654.g003

Genetic Analysis

SNP genotyping and quality control. Genomic DNA was extracted from 7 ml of EDTA-treated blood with standard methods. SNPs genotyping for all samples was performed using the Affymetrix GeneChip platform. We utilized the GeneChip® Human Mapping 500 K Array Set that comprises two arrays (the Nsp and Sty arrays) capable of genotyping ~262000 and ~238000 SNPs, respectively. We followed the recommended protocol described in the Affymetrix manual. All DNA samples were normalized to 50 ng/μl. Then, 5 μl (250 ng) of dsDNA was digested with the appropriate restriction enzyme and ligated to adapters using T4 DNA ligase. Samples were then PCR amplified using TITANIUM Taq polymerase on an GeneAmp® PCR System 9700 gold plate thermal cycler. PCR products were purified using the Clontech purification kit followed by fragmentation. Samples were then injected into cartridges, hybridized, washed and stained. Mapping array images were obtained using the GeneChip Scanner 3000 7G plus. For quality control (QC), individual arrays not passing the 93% call rate threshold at $P = 0.33$ with the Dynamic Model algorithm [36] were considered a failure and re-genotyped. The average QC call rate of the data produced using the Nsp arrays was 95.25% while the Sty array reached to 95.56%. All individuals passing the QC quality checks were further screened based on genotyping performance. Genotypes were called using the BRLMM (Bayesian Robust Linear Model with Mahalanobis distance classifier) software [37]. The average call rate of entire sample was 98.66% and 98.82% for Nsp and Sty fractions respectively. Any individual with genotyping call rates less than 95% for either Nsp or Sty fractions were excluded. All individuals with call rates >95% showed an average of 91.07% SNPs call rate. In addition, individuals whose gender call from X chromosome genotype data was discrepant with the gender obtained from medical records were excluded from the analysis.

All SNPs physical coordinates refer to the NCBI released annotation update for the human genome (NCBI Build 36.2). Alleles are expressed in the forward (+) strand of the reference.

The mean minor allele frequencies (MAF) in our isolated population using the 500 K Array GeneChip® showed an average of 0.187 (± 0.003). After the exclusion of monomorphic SNPs, the MAF increased an average of 0.251 (± 0.004). We selected only SNPs informative (MAF > 0.01) in all villages obtaining a sub-sample of 361980 SNPs (average space 7672 bp).

Linkage Disequilibrium Maps and Pair-wise LD metrics. LDMap program [38] was used to construct LD maps. The LD map describes the fine variations in LD pattern over a given chromosomal segment, calculating linkage disequilibrium units (LDU) between adjacent pairs of SNPs. The LDU scale is constructed from the product of physical (kb) distance and a parameter describing the exponential decline in association with distance computed for each interval between adjacent SNPs. In the LD maps, the regions with high LD can be identified as plateaus where the increase in LDU is very small or zero, while the regions with lower LD can be identified as steps. The LD maps were calculated on autosomes based on all 361980 SNPs for the 360 samples. Thus, comparison of LD maps in the villages was based on the same SNPs and on the same number of samples.

Two types of pair-wise LD metrics were calculated: Lewontin's standardized deviation coefficient $|D'|$ [39] and the pair-wise correlation r^2 [40] between all pairs of SNPs within 500 Kb of each other, using the program Haploview [41] (<http://www.broad.mit.edu/haploview/>).

Structure and population analysis

Single nucleotide polymorphisms (SNPs) analysis. For population analysis, 500 K were filtered according to inter-SNP

distance (about 500 kb) to minimize LD effects, obtaining 5,262 SNPs. We tested Hardy-Weinberg equilibrium with a modified version of the Markov-chain random walk algorithm described in Guo and Thompson [42]. The modified version gives the same results as the original one, but is more efficient from a computational point of view. The analysis was performed with the software Arlequin 3.1 [43], (<http://cmpg.unibe.ch/software/arlequin3/>). Seventy markers were excluded from the following analysis, as they showed highly significant departure from Hardy Weinberg equilibrium ($P < 10^{-5}$), obtaining 5,192 markers for the following analysis.

To prepare haplotype data, we considered 83 genomic regions with $D' > 0.7$ in all 8 communities, starting from data employed for LD analysis, for a total of 361980 SNPs. Since recombination in X-linked and autosomal portions of the human genome is often concentrated in hotspots separated by DNA regions with little or no recombination [44,45], we felt that our choice of high LD genomic regions (precisely $D' > 0.7$) in our samples would be the most informative.

We estimated haplotype frequencies using windows of 10 consecutive informative SNPs for each region, spanning an average of 150 kb.

Inference of haplotype frequencies were performed employing the software PHASE 2.1 [46,47].

Haplotype data were used to compute genetic distances. Each region was considered as a single polymorphic locus, where each haplotype corresponded to a single allele.

F_{ST} , a standard measure of genetic differentiation between populations [48] was estimated with the Weir and Cockerham algorithm [49]. F_{ST} is the proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance. Values can range from 0 to 1, and a high value implies a considerable degree of differentiation among populations. The 95% confidence limits for F_{ST} were determined by 1000 permutations test. Level of statistical significance was tested by performing 1000 permutations. Both analysis were performed with GENETIX 4.5.02 software [50], (<http://www.genetix.univmontp2.fr/genetix/genetix.htm>).

To ascertain the proportions of genetic variance due to differences within and between populations, genetic variance was hierarchically apportioned according to geographic criteria through the locus by locus AMOVA (analysis of molecular variance) [51], using the Arlequin 3.01 software [43]. Number of permutations was set at 10000.

Genetic distances were computed from allele and haplotype frequencies according to Nei [52] and Reynolds [53], with the PHYLIP v 3.66 [54] executable "Gendist", after processing of dataset for bootstrapping (1000 replicates), using PHYLIP executable "Seqboot". We applied two different measures of genetic distances to test different assumptions. All assume that all differences between populations arise from genetic drift. Nei's [52] distance model, is based on infinite isoallele model of mutations [55], where all loci have the same rate of neutral mutations, and that the genetic variability originally in the population is at equilibrium between mutations and genetic drift. Moreover, the effective population size of each population remains constant. Reynolds' distance [53] assumes that there is no mutation, so that gene frequency changes are by genetic drift alone. In addition, population sizes do not remained constant and equal in all populations.

Distances were clustered using the Neighbor - Joining method [56], and consensus tree were built with the PHYLIP executable "Consense". We used the Neighbor Joining method instead of other methods as UPGMA (Unweighted Pair Group Method of Analysis) or ML (Maximun Likelihood) because it was shown to be generally better than the other methods [56]. Phylip package is

available at the following link: <http://evolution.genetics.washington.edu/phylip.html>.

We performed factor Correspondence Analysis (FCA) using the Genetix 4.5.02 software [50]. FCA [57–59] is a kind of canonical analysis particularly well suited to describe associations between two qualitative variables (analysis of a contingency table crossing the terms of two variables). She et al. [60] have proposed a more appropriate coding of data for the genetics of diploid organisms, whereas others authors [61–63] define the correlation between the results of the analysis and conventional parameters of population genetics.

Details of mathematical principles of methods are provided at the following link: http://www.unesco.org/webworld/idams/advguide/Chapt6_5.htm. We used this method because Guinand [63] indicated that CRT-MTC (whose FCA implemented in Genetix utilize a similar strategy) presents various advantages in respect to PCA (Principal Component Analysis) in estimating some parameters and in the presence of a pooling strategy. Correspondence analysis has been used recently with good results in the Seldin et al. study [8].

We examined correlation between genetic and geographic distance matrices using the Mantel test by a permutation procedure. The permutation allows to examine the empirical null distribution of the correlation coefficient taking into account the auto-correlations of the elements of the matrix [64,65]. Here, we compared genetic distance computed with the described method obtaining a null distribution with 500000 permutations. The analysis was performed with Arlequin 3.1 software [43].

To infer individual ancestry and population admixture, we carried out analysis of population structure using a model-based clustering method developed in the STRUCTURE 2.2 software [66] (<http://pritch.bsd.uchicago.edu/structure.html>). The model assumes the presence of K populations (where K may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned probabilistically to one cluster or to two or more clusters if their genotypes indicate that they are admixed [67]. We conducted the analysis according to the admixture model and without any prior population assignment. We performed several runs for each K (from 2 to 8), until we obtained 10 times the same result, using 20000 replicates and 10000 burn-in cycles.

Results

Measuring Linkage Disequilibrium

We compared the extent of LD for the 8 villages and for a sample originating from other geographical areas of Sardinia. We used a set comprising both commons and informative SNPs in all communities. Moreover, to compare our results to a standard outbred population, we included in our study 60 unrelated samples from 30 HapMap CEU trios. The mean heterozygosity of the markers was 0.345 (SD = 0.011) in the 10 communities selected, ranging from 0.320 (Urzulei) to 0.361 (Seui). The average of minor allele frequency (MAF) was 0.251 (SD = 0.004), ranging from 0.241 (Urzulei) to 0.257 (Seui).

The monomorphic SNPs (M-SNPs), in at least one village, were ~26% of the entire SNPs list. The shared M-SNPs in the eight villages were 31656. The comparison with the HapMap samples showed that 88.5% of the shared M-SNPs were monomorphic also in the CEU population and 10.89% showed a $MAF > 0$ and < 0.05 . Considering the total number of M-SNPs in each village, we identified a group of village specific M-SNPs that range from the highest values in Urzulei and Talana (5.04% and 4.33% respectively) to the lowest in Seui and Loceri (1.85% and 1.45% respectively).

We constructed LD maps for the 10 populations on autosomes. The shortest maps were in Talana, Urzulei, Ussassai, Trieri and Seulo, in agreement with their small population size and their secluded position. We found the trend growing from Baunei to Seui and to Loceri. The longest maps were in the Sardinia and in the CEU samples (figure 4). Similar results and differences in LD maps were highlighted in all chromosomes analyzed. Moreover the metric LD map evidence a pattern of “plateaus”, regions of high LD, and “steps”, regions of increased recombination: this pattern tends to be shared among all villages. In addition, the comparison with CEU population confirmed this pattern and presumably reflects the common distribution of recombination hot spots across various populations [68].

We compared the extent of LD by examining the distribution of D' and r^2 on autosomes in the selected communities. We calculated D' and r^2 between all pairs of markers (figure 5). The highest D' and r^2 average values were detected in Talana, Urzulei, Ussassai, Trieri and Seulo, and the lowest in Sardinia and CEU.

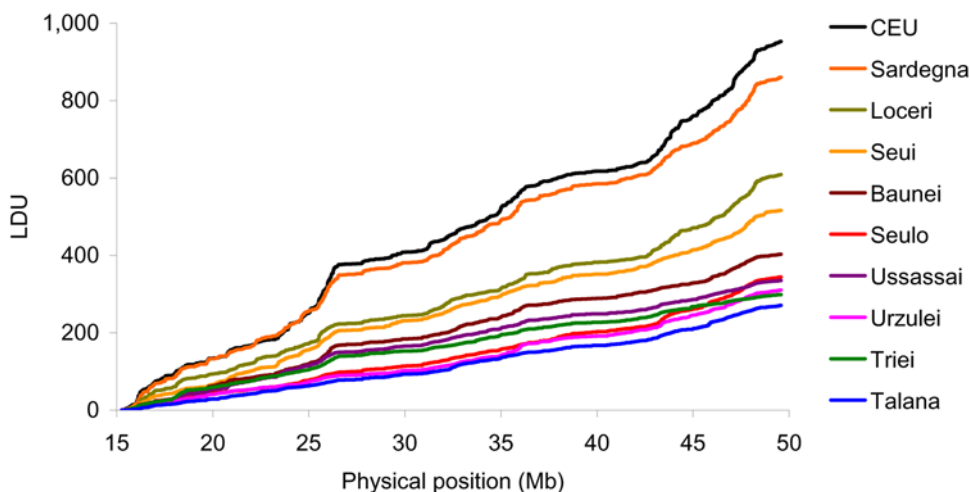


Figure 4. Comparison of population-specific LD maps on chromosome 22 of the 10 samples from Ogliastra villages, rest of Sardinia and CEU.

doi:10.1371/journal.pone.0004654.g004

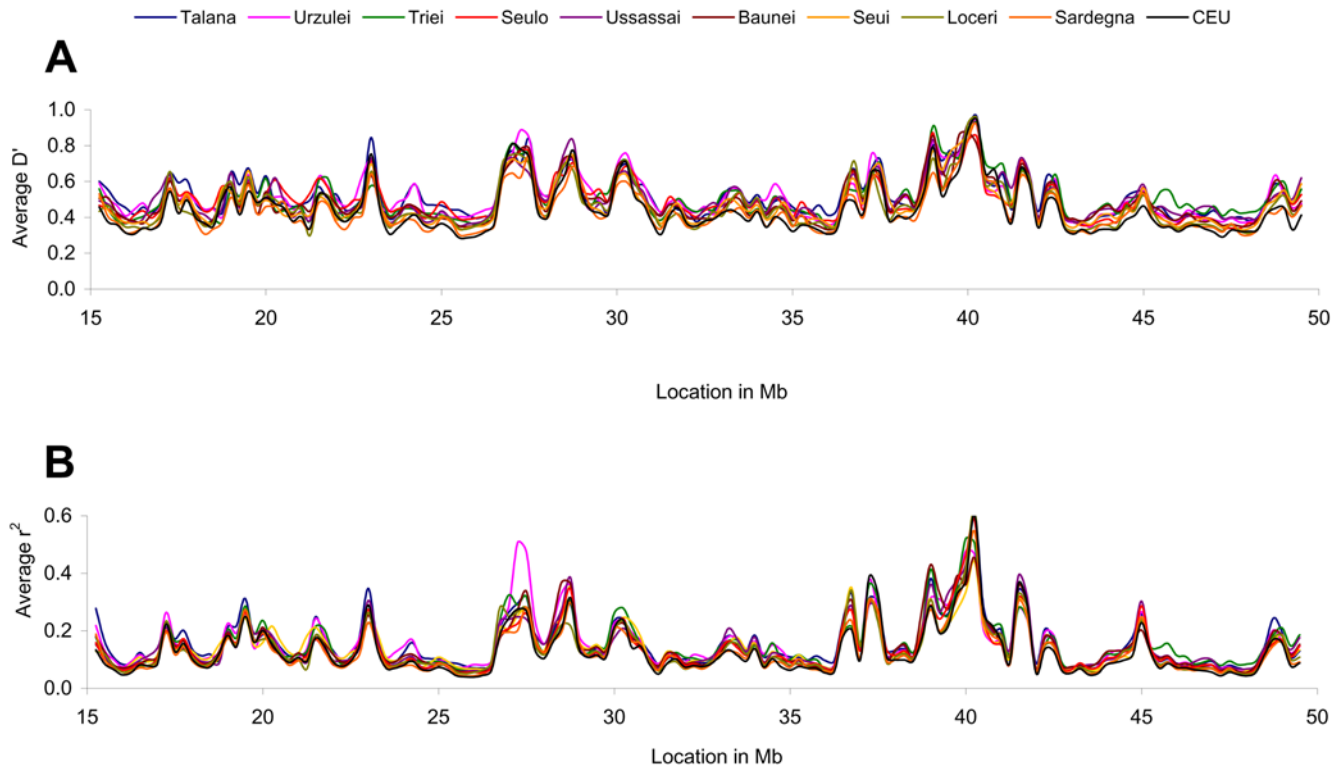


Figure 5. Distribution of Linkage Disequilibrium on chromosome 22. Average D' (A) and r^2 (B) coefficients plotted in 0.5 Megabases sliding windows (0.25 Mb overlap).
doi:10.1371/journal.pone.0004654.g005

In performing LD analysis in small populations it is important to consider the kinship value among the subjects, because most individuals are in some way related to each other. This relatedness may inflate estimates of LD. Utilizing the genealogical information present in our database, we reconstructed all the family relations for each individual in each villages and calculated the pair wise relatedness of individuals. We selected 45 unrelated individuals from each village reaching a total of 360 subjects (table 1). The average kinship value goes from the highest value in Talana (0.021) to the lowest in Loceri and Baunei (0.004). To evaluate how the degree of relatedness influences LD, we analyzed the LD

pattern on 10 randomly chosen groups of 45 individuals each from Talana. The average kinship of these 10 groups is comprised between 0.022 and 0.025. The LD map length of the randomly chosen groups ranged from 292 to 288 LDU.

Population structure analysis

In order to assess the level of differentiation among the villages we calculated the F_{ST} values by the Weir and Cockerham [49] algorithm (table 2). F_{ST} for the 8 communities was 0.0213. We found the highest difference between the villages of Talana and Seulo ($F_{ST} = 0.0324$) and the lowest difference between Baunei and Triei ($F_{ST} = 0.0105$). All the comparisons between villages were highly significant ($P < 0.0001$).

For each village, we computed average F_{ST} resulting from each pairwise comparison. We observed the highest degree of differentiation for Talana and Seulo villages, with $F_{ST} = 0.0260$ and $F_{ST} = 0.0238$, respectively. Moreover, these two villages show a great deal of genetic distance between each other ($F_{ST} = 0.0324$).

When we computed F_{ST} values pooling villages according to geographic position (Talana, Urzulei, Baunei, Triei in the first group and the others villages in the second), we observed a higher F_{ST} value in the first group ($F_{ST} = 0.0200$) than in the second ($F_{ST} = 0.0142$).

For AMOVA analysis [51], we grouped villages in two groups according to geographic position, as in the previous F_{ST} analysis. Results reveal a statistically significant genetic heterogeneity between the two geographic areas (F_{CT}) (0.75% of total variation, $P < 10^{-5}$). Also the heterogeneity among villages within areas (F_{SC}) and among individuals within villages (F_{ST}) is highly significant (1.69%, $P < 10^{-5}$ and 97.56%, $P < 10^{-5}$).

Population structure was analyzed with the mean of the Mantel Test, which compares coefficient correlation between geographic and

Table 1. For each village we report resident population size, average village kinship values, number of samples and its average kinship values.

	Today population	Average Village kinship	Samples	Average Sample Kinship
Talana	1129	0.024	45	0.021
Urzulei	1443	0.017	45	0.013
Baunei	3886	0.006	45	0.004
Triei	1115	0.020	45	0.016
Seulo	1023	0.016	45	0.014
Seui	1587	0.007	45	0.005
Ussassai	763	0.017	45	0.012
Loceri	1336	0.010	45	0.004

doi:10.1371/journal.pone.0004654.t001

Table 2. Fst values computed for each pair-wise comparison calculated on 5262 SNPs evenly spaced on 500 Kb. Corresponding 95% confidence intervals, shown between parentheses, were determined with permutations testing (set at 1000).

	Baunei	Loceri	Seui	Seulo	Talana	Triei	Urzulei
Loceri	0.01563 (0.01462–0.01664)						
Seui	0.01970 (0.01843–0.02096)	0.01089 (0.01007–0.01172)					
Seulo	0.02520 (0.02396–0.02649)	0.01724 (0.01628–0.01836)	0.01435 (0.01335–0.01536)				
Talana	0.02296 (0.02166–0.02442)	0.02197 (0.02081–0.02319)	0.02590 (0.02447–0.02727)	0.03238 (0.03061–0.03404)			
Triei	0.01051 (0.00971–0.01136)	0.01854 (0.01737–0.01971)	0.02245 (0.02112–0.02371)	0.02883 (0.02734–0.03033)	0.02500 (0.02361–0.02643)		
Urzulei	0.01719 (0.01607–0.01831)	0.01769 (0.01651–0.01871)	0.02312 (0.02172–0.02444)	0.02942 (0.02790–0.03090)	0.02346 (0.02213–0.02482)	0.02051 (0.01923–0.02174)	
Ussassai	0.02524 (0.02395–0.02664)	0.01435 (0.01335–0.01536)	0.01096 (0.01009–0.01187)	0.01938 (0.01813–0.02060)	0.03051 (0.02894–0.03210)	0.02766 (0.02616–0.02923)	0.02728 (0.02577–0.02863)

The levels of statistical significance were tested by performing 1000 permutations. All comparisons were highly significant ($P < 10^{-3}$). doi:10.1371/journal.pone.0004654.t002

genetic distances [64,65]. Also in this case, we performed the analysis using both Nei [52] and Reynolds [53] distances. We observed in both SNPs and haplotype analysis, a significant correlation ($P = 0.004$ and $P = 0.001$), indicating a pattern of isolation by distance in the Ogliastro genetic pool (figure 6A and 6B).

Figure 7 shows the Neighbor Joining tree computed with Reynolds [53] genetic distances from allele (7A) and haplotype frequencies (7B). We performed the same analysis with Nei [52] distances, with identical results (data not shown). The tree topology corresponds to the geographic distribution of the two cluster areas in the Ogliastro region with the village of Loceri in an intermediate position.

We plotted results of Factor Correspondence Analysis [63] performed with allele frequencies of the 8 communities for factors 1 and 2 (26.5% and 17.6% of inertia, respectively) and for factors 1 and 3 (26.5% and 15.4%, respectively). Results confirm differentiation between the two geographic areas (figure 8A1–8A2), separated by factor 1, highlight the strong differentiation of Talana-Urzulei versus other villages (factor 2), and of Talana versus Urzulei (factor 3). We repeated the computation excluding Talana and Urzulei, to better visualize patterns of population relationships among the remaining six villages, (figure 8B1–8B2). Factor 1 accounts for 35.9% of inertia, and factor 2 and 3 represented 20.8% and 15.9% of inertia. This analysis confirms the differentiation in two main regions, separated by factor 1. Factor 2 highlights the intermediate position of Seui between Seulo and Ussassai, and factor 3 suggests a notable differentiation between Baunei and Triei.

To performed STRUCTURE analysis [66], we assessed number of population groups (K), performing 10 runs at each K , from 2 to 8. The estimation of log_e probability of the data using the F model favored the assumption of $K = 7$. Figure 9 shows results at different K values. Analysis at $K = 2$ illustrates the differences of the two regions, while at subsequent K values we noted the differentiation of Talana, Urzulei, Baunei and Triei in comparison with southern Ogliastro. The remaining villages are included at subsequent K values. At $K = 8$, we detect 8 different clusters which do not correspond to each single village: Triei and Baunei were in the same cluster while Loceri is composed by two different clusters.

Discussion

We analyzed features and different aspects of population structure in eight villages of Ogliastro region in the central-eastern area of Sardinia. This region is characterized by high endogamy, low immigration, environmental homogeneity, as well as genetic differentiation from the rest of the island [21]. Gene diversity measures of HVSI mtDNA haplotypes suggest that Ogliastro ranks among the most genetically homogeneous European populations [23]. Furthermore, Ogliastro has lowest values of mtDNA gene diversity respect to other areas of Sardinia [23,27,69]. A more precise analysis of distinct sub-populations in the same region, revealed a striking differentiation due to distinctive founder effects and genetic drift. This was confirmed by our studies of mtDNA and Y chromosome haplotypes on few Ogliastro villages [27,32,70]. Such differentiation was hinted at by our previous LD analysis with microsatellite markers in the Xq13.3 region [17].

In this paper we focus our analysis on eight isolated Ogliastro villages to evaluate LD parameters and effective population structure using a dense whole genome SNPs map. The length of LD map is inversely related to the extent of LD over a given chromosomal segment, therefore shorter LD maps are observed in population isolates compared with more heterogeneous popula-

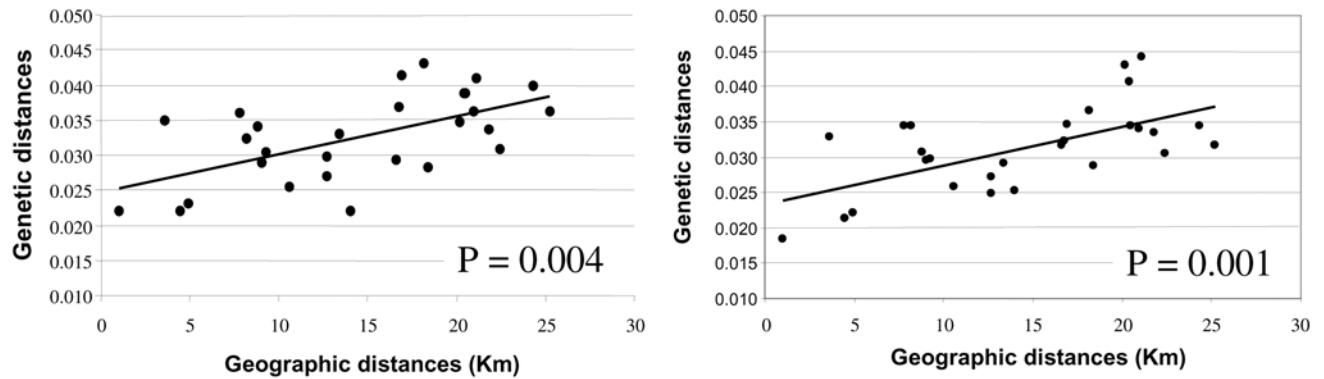


Figure 6. Genetic distance as a function of geographic distance between each pair of communities. On the left (A) Mantel Test plot analysis using SNPs data; on the right (B) Mantel Test plot analysis using Haplotype data.
doi:10.1371/journal.pone.0004654.g006

tions. Although the geographic area analyzed in this study was relatively restricted, our results evidence high values of LD parameters and the presence of LD variability across the genome in these communities depending on their historical demography and population structure. The 5 smallest and secluded villages show the shortest LD maps, in agreement with their similar size and isolated condition: from 270 LDU in Talana to 343 LDU in Seulo. We identified intermediate values in Baunei (LDU = 402), a large community with a continuous growth from the 17th century to today.

Seui, on the other hand, has peculiar historical and demographic aspects. This village, which was already an important administrative centre during the Spanish domination of Sardinia (even has a prison dating from the 1640s) has undergone immigration waves starting from the mid-19th century due to the presence of anthracite and copper mines. During the 19th century, Seui became one of Ogliastra main economical centre and the population grew to about 3000 inhabitants from the beginning of the 20th century to the 1960s. Then there was a massive emigration wave toward the main political and economic centre

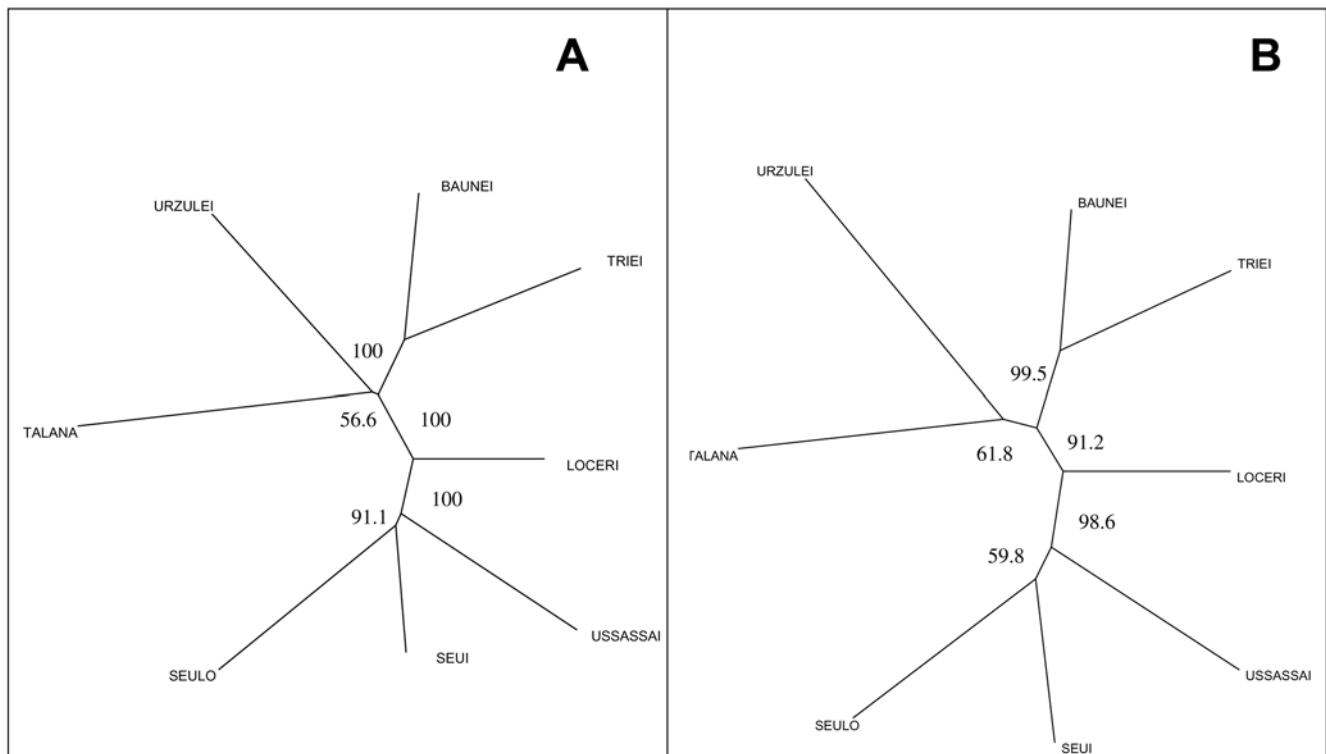


Figure 7. Neighbor Joining trees analysis. A) Neighbor Joining tree computed with Reynolds (1983) distances from allele frequencies of 5192 SNPs. Bootstrap supports over 50% (out of 1000 iterations) along the nodes. B): Neighbor Joining tree computed with Reynolds (1983) distances from haplotype frequencies of 83 regions containing 10 SNPs. Bootstrap supports over 50% (out of 1000 iterations) along the nodes.
doi:10.1371/journal.pone.0004654.g007

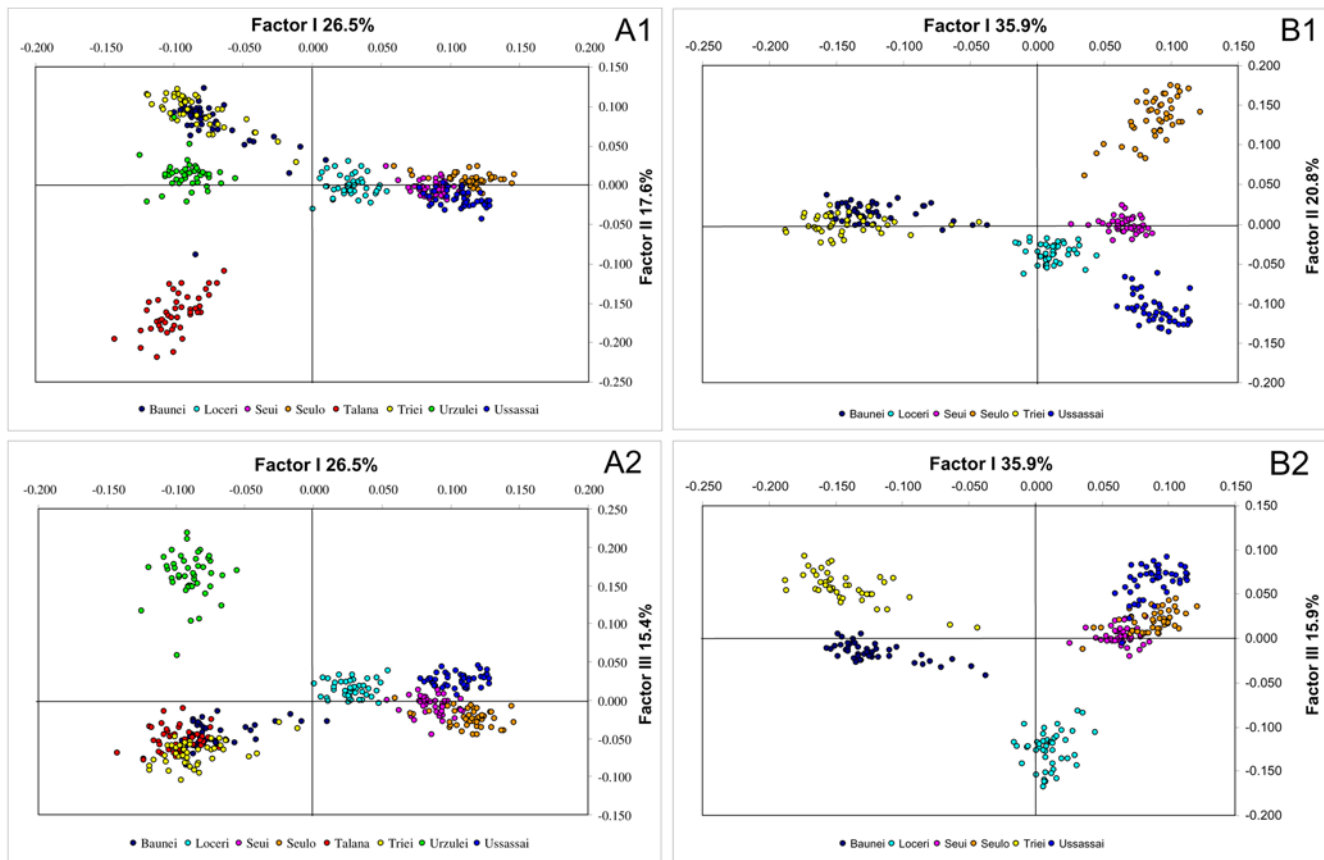


Figure 8. Factor Correspondence Analysis comparing different individuals from different Ogliastra villages, performed with 5,192 SNPs. Each individual is represented by a circular shape, and the different 8 communities are marked with different colours. The two plots represented factor 1 and 2 and factor 1 and 3. (A1, A2) Analysis performed with 8 communities; (B1, B2) Analysis performed with 6 communities, without Talana and Urzulei.

doi:10.1371/journal.pone.0004654.g008

of Sardinia (Cagliari) with a rapid decrease of resident population. Today there are only 1587 inhabitants, and this has influenced LD parameters and current population structure.

Finally, Loceri has the lowest LD values because it is not in an isolated geographical position, it is close to the seashore and to Lanusei, an important Ogliastra administrative centre.

Its frequent genetic exchanges with other villages are proven by archival data spanning many centuries.

Additionally, the extreme features of some villages such as Talana and Urzulei, were evidenced by the analysis of the monomorphic SNPs. The different percentage of specific monomorphic SNPs shows a trend similar to that of different populations, it is consistent with our analyses of LD values and it contributes to evidence the different level of isolation of these villages.

The STRUCTURE population analysis confirms the presence of a distinct cluster for each village analyzed, except for Baunei and Triei (which in the past were part of a single municipality). Moreover, we observed high values of F_{ST} among villages. Recently various studies have analyzed comparable set of data obtained from European and East Asian Populations [71,72]. We compared our results with Europeans. In the analysis of Italian, Swedish and Spanish cohorts [8], the total value of F_{ST} from a subset of 5,700 SNPs distributed over entire genome, computed with the same method ($F_{ST} = 0.0029$) was lower than our sample ($F_{ST} = 0.0213$). The highest degree of differentiation has been

detected between samples from Italy and Sweden ($F_{ST} = 0.0060$). This value is lower than the one detected between Baunei and Triei ($F_{ST} = 0.0105$), which are the least differentiated villages in this study. Salmela et al. [72] analyzed the structure in various North Europe populations (Finland, Sweden, Northern Germany and Great Britain) and computed F_{ST} values in a subset of 6369 SNPs, finding values lower (from 0.0005 to 0.0072) than values observed among Ogliastra villages (0.0105 to 0.0324).

Our data suggest a great level of differentiation among Ogliastra villages: even if this result is influenced by the difficulty in detecting micro differentiation in countrywide samples, nevertheless it clearly reveals the high level of population substructure in Ogliastra.

In general, our results show a good correlation with the geographical structure of the region, as revealed by the pattern of isolation by distance of the Mantel Test. In particular, Neighbor Joining trees computed from different sets of data (single SNPs and haplotypes) confirmed the Mantel Test results because villages are located in the tree according to their geographic position and relative closeness. Furthermore, the topology of the tree shows two main clusters identifying a Northern and a Southern Ogliastra sub-area. This is confirmed by AMOVA and STRUCTURE analysis at low level of differentiation ($K = 2$), highlighting the separation of the two sub-areas.

Factor Correspondence Analysis and STRUCTURE analysis emphasize that Talana and Urzulei are highly differentiated

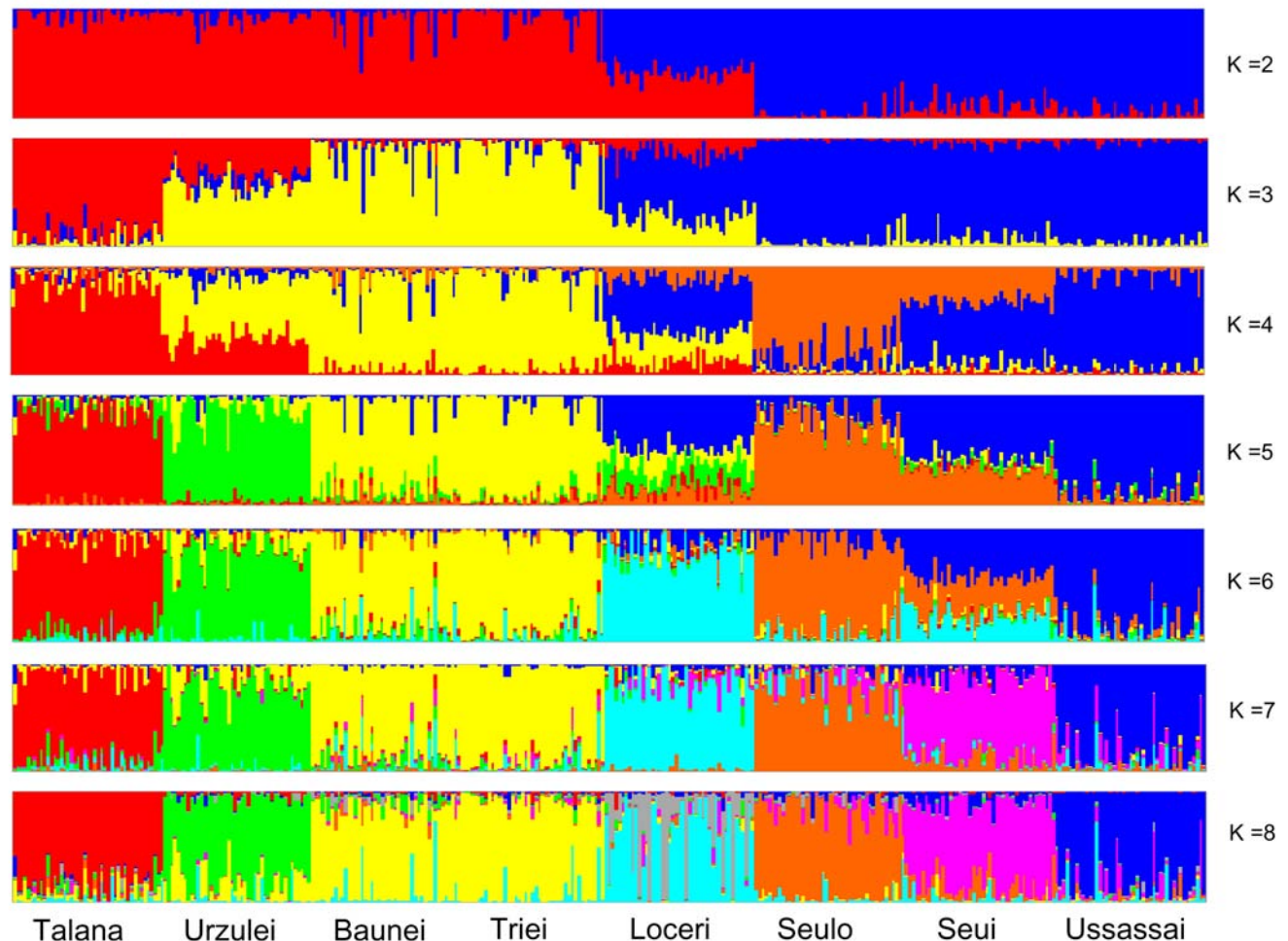


Figure 9. STRUCTURE results under the assumption of different population groups (K = 2:8). Each individual is shown as a vertical line partitioned into K colored components representing inferred membership in K genetic clusters. K = 7 represents the highest values of $\ln P(D)$. doi:10.1371/journal.pone.0004654.g009

although separated by limited geographic distance. In addition, Neighbor Joining tree shows the village of Loceri to be in an intermediate position, while correspondence analysis confirms that Triei and Baunei are highly correlated but not identical.

Our results, obtained with a large SNPs data set, confirm a high degree of differentiation among villages, leading us to some considerations: not only small areas as Ogliastra should not be considered homogeneous a priori, but even adjacent villages could be not homogeneous.

A number of analyses performed in the large and small genetic isolates showed either the presence or absence of substructure. For example, in Hutterites communities we cannot identify distinct subpopulation [73]. Alternatively, the analysis of the Icelandic genetic pool, carried out with 40 microsatellite markers in 43748 individuals from 11 different geographic regions of the island, showed the presence of notable regional subdivisions. Icelanders cannot be considered to be a single randomly interbreeding population [7]. Similarly, the analysis of 14 biallelic markers and 8 STRs of Y-chromosome on samples from different areas of Finland revealed the existence of a sharp genetic border between eastern and western Finland [11], confirmed also by epidemiological differences [74,75]. Furthermore, a recent analysis with genome wide SNP data, demonstrate the presence of substructure among 10 distinct Finnish early and late-settlement subpopulations [10].

In the case of small population isolates, the analysis of 6 South Tyrolean villages performed with microsatellites on Xq13.3, showed the presence of three genetically distinct sub-populations, sharing the alpine environment and lifestyles [76]. Vitart et al. [77] found genetic differentiation among 10 small isolated villages in Croatia islands. Estimated F_{ST} values were 0.02 (95% CI: 0.017–0.022), with many significant differences in the comparison of these villages.

The information obtained on the genetic differences present among our villages in Ogliastra is potentially useful for the ongoing extensive genetic research in this specific population. Detailed knowledge of the high differentiation in the degree of LD background found in each village will probably be crucial in the analysis of these sub-isolates that represent useful populations for the initial detection of loci/genes with Genome-wide association (GWA) involved in the predisposition to complex traits. Such analysis require a great deal of attention to exclude false positive results as consequence of stratification differences between cases and controls [78–80]. Association studies can be confounded by differences in ancestry: geographic ancestry can explain just a portion of human genetic variation. Several genomic variants had high frequencies in populations e/o cohort of samples with particular ancestries and such variants could erroneously appear to be related to disease. These incorrect results can be attributed in

part to complex relationships reflecting different population origins that included migration, admixture and isolation.

There are potential advantages in the choice of these populations because of reduced environmental complexity and a probably reduced number of disease alleles. Some of the loci are expected to have stronger genetic effects on the disease/trait under analysis. Furthermore, small isolates might contain mutations *e/o* possible variants that are rare in the population at large, thus allowing the identification of genes that would otherwise be missed [81].

One negative aspect could be the limited number of cases in association studies, except for common diseases with high prevalence. In fact our group was able to identify a variant related to the EDA2R gene strongly associated with Androgenetic Alopecia showing an average prevalence of 47% in the 8 villages [82].

We also propose that this kind of populations is helpful in the search for variants associated to quantitative traits (QT) related to common disease. In QT loci analyses all individuals phenotyped are useful and even the limited number of inhabitants of one of our villages could be a sufficient cohort for initial association studies. Once a significant association is obtained we can replicate the findings in the other villages. Obviously, it will be important to replicate the findings in a general Sardinian sample and in outbred populations.

References

- Bonnen PE, Peter I, Plenge RM, Salit J, Lowe JK, et al. (2006) Evaluating potential for whole - genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 38: 214–217.
- Bourgain C (2005) Comparing strategies for association mapping in samples with related individuals. *BMC Genet* 6: S1–S98.
- Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. *Nat Genet* 23: 397–404.
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, et al. (2004) Assessing the impact of population stratification on genetic association studies. *Nat Genet* 36: 388–393.
- Clayton D (2001) Population association. In: Balding DJ, Bishop M, Cannings C, eds. *Handbook of Statistical genetics*. Chichester, UK: John Wiley and Sons. pp 519–540.
- Marchini J, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. *Nat Genet* 36: 512–517.
- Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. *Nat Genet* 37: 90–95.
- Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, et al. (2006) European population substructure: clustering of northern and southern populations. *Plos Genet* 2: e143.
- Tian C, Plenge RM, Ransom M, Lee A, Villoslada, et al. (2008) Analysis and application of European genetic substructure using 300 K SNP information. *Plos Genet* 4: e4.
- Jakkula E, Rehnström K, Varilo T, Pietiläinen OP, Paunio T, et al. (2008) The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet* 83: 787–794.
- Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, et al. (2006) Regional differences among the Finns: a Y-chromosomal perspective. *Gene* 376: 207–215.
- Feder J, Ovadia O, Glaser B, Mishmar D (2007) Ashkenazi Jewish mtDNA haplogroup distribution varies among distinct subpopulations: lessons of population substructure in a closed group. *Eur J Hum Genet* 15: 498–500.
- Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. *Nat Rev Genet* 5: 89–100.
- International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 31 million SNPs. *Nature* 449: 851–861.
- Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, et al. (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38: 56–60.
- Lee N, Daly MJ, Delmonte T, Lander ES, Xu F, et al. (2001) A Genomewide Linkage Disequilibrium scan localizes the Saguenay–Lac-Saint-Jean Cytochrome Oxidase Deficiency to 2p16. *Am J Hum Genet* 68: 397–409.
- Angius A, Bebbere D, Petretto E, Falchi M, Forabosco P, et al. (2002) Not all isolates are equal: linkage disequilibrium analysis on Xq133 reveals different patterns in Sardinian sub-populations. *Hum Genet* 111: 9–15.
- Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, et al. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 9: 2947–2957.
- Rosatelli MC, Dozy A, Faà V, Meloni A, Sardu R, et al. (1992) Molecular characterization of beta-thalassemia in the Sardinian population. *Am J Hum Genet* 50: 422–446.
- Loudianos G, Dessi V, Lovicu M, Angius A, Altuntas B, et al. (1999) Mutation analysis in patients of Mediterranean descent with Wilson disease: identification of 19 novel mutations. *J Med Genet* 36: 833–836.
- Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, et al. (1996) Genetic analysis of Sardinia: I data on 12 polymorphisms in 21 linguistic domains. *Ann Hum Genet* 60: 125–141.
- Latini V, Vacca L, Ristaldi MS, Marongiu MF, Memmi M, et al. (2003) Beta-globin gene cluster haplotypes in the Corsican and Sardinian populations. *Hum Biol* 75: 855–871.
- Fraumene C, Petretto E, Angius A, Pirastu M (2003) Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum Genet* 114: 1–10.
- Vona G (1997) The peopling of Sardinia (Italy): history and effects. *Int J Anthropol* 12: 71–87.
- Morelli L, Grosso MG, Vona G, Varesi L, Torroni A, Francalacci P (2000) Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum Biol* 72: 585–595.
- Lampis R, Morelli L, De Virgiliis S, Congia M, Cucca F (2000) The distribution of HLA class II haplotypes reveals that the Sardinian population is genetically differentiated from the other Caucasian populations. *Tissue Antigens* 56: 515–521.
- Ghiani ME, Vona G (2002) Y-chromosome-specific microsatellite variation in a population sample from Sardinia (Italy). *Coll Antropol* 26: 387–401.
- Zei G, Lisa A, Fiorani O, Magri C, Quintana-Murci L, et al. (2003) From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *Eur J Hum Genet* 11: 802–807.
- Contu D, Morelli L, Santoni F, Foster JW, Francalacci P, et al. (2008) Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans. *Plos One* 3: e1430.
- Cavalli-Sforza LL, Menozzi P, Piazza A (1996) *The history and geography of human genes*. Princeton, New Jersey: Princeton University Press. 413 p.
- Cocco F (1987) *Dati relativi alla Storia dei paesi della diocesi d'Ogliastra (Vol III)*. Cagliari, Italy: Tipografia TEA. 306 p.
- Angius A, Melis PM, Morelli L, Petretto E, Casu G, et al. (2001) Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. *Hum Genet* 109: 198–209.
- Angius A, Hyland FC, Persico I, Pirastu N, Woodage T, et al. (2008) Patterns of Linkage Disequilibrium between SNPs in a Sardinian Population Isolate and the Selection of Markers for Association Studies. *Hum Hered* 65: 9–22.
- Relethford JH, Mielke JH (1994) Marital exogamy in the Aland Island, Finland, 1750–1949. *Ann Hum Biol* 21: 13–21.
- Wright S (1922) Coefficients of inbreeding and relationship. *Am Nat* 56: 330–338.

36. Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, et al. (2005) Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* 21: 1958–1963.
37. Rabbee N, Speed T (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7–12.
38. Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. *Am J Hum Genet* 54: 705–714.
39. Lewontin RC (1964) The interaction of selection and linkage I General considerations; heterotic models. *Genetics* 49: 49–67.
40. Maniatis N, Collins A, Xu C-F, McCarthy LC, Hewett DR, et al. (2002) The first linkage disequilibrium (LD) maps: Delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci USA* 99: 2228–2233.
41. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
42. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.
43. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinf Online* 1: 47–50.
44. Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, et al. (2000) Recombinational and mutational hotspots within the human *Lipoprotein Lipase* gene. *Am J Hum Genet* 66: 674–686.
45. McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, et al. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581–584.
46. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *Am J Hum Genet* 76: 449–462.
47. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978–989.
48. Wright S (1978) *Evolution and the Genetics of Populations* (Vol. 4): Variability within and among natural populations. Chicago: University of Chicago Press.
49. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
50. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2003) GENETIX, logiciel sous Windows™ pour la génétique des populations Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
51. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
52. Nei M (1972) Genetic distances between populations. *Amer Nature* 106: 283–292.
53. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short term genetics distance. *Genetics* 105: 767–779.
54. Felsenstein J (1989) PHYLIP: Phylogeny Inference Package (version 32). *Cladistics* 5: 164–166.
55. Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite Population. *Genetics* 49: 725–738.
56. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
57. Benzécri JP (1973) *L'Analyse des Données: T. 2, P Analyse des correspondances*. Paris: Dunod.
58. Greenacre MJ (1984) *Theory and applications of correspondence analysis*. London: Academic Press.
59. Escoufier B, Pagès J (1990) *Analyses factorielles simples et multiples*. Paris: Dunod.
60. She J-X, Autem M, Kotoulas G, Pasteur N, Bonhomme F (1987) Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biol J Linn Soc* 32: 357–371.
61. Long JC (1986) The allelic correlation of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's F-Statistics. *Genetics* 112: 629–647.
62. Smouse PE, Long JC (1988) A comparative F-statistics analysis of the genetic structure of human populations from Lowland South America and Highland New Guinea. In: Weir BS, Eisen G, Goodman MM, Namkoong G, eds. *Second international conference in quantitative genetics*. Sunderland (MA): Sinauer Associates. pp 32–46.
63. Guinand B (1996) Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biol J Linn Soc Lond* 58: 173–195.
64. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27: 209–220.
65. Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel Test of matrix correspondence. *Sys Zool* 35: 627–632.
66. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*.
67. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
68. De La Vega FM, Isaac H, Collins A, Scafe CR, Halldórsson BV, et al. (2005) The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res* 15: 454–462.
69. Falchi A, Giovannoni L, Calò CM, Piras IS, Moral P, et al. (2006) Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. *J Hum Genet* 51: 9–14.
70. Fraumene C, Belle EM, Castri L, Sanna S, Mancosu G, et al. (2006) High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. *Mol Biol Evol* 23: 2101–2111.
71. Salmela E, Lappalainen T, Fransson I, Andersen PM, Dahlman-Wright K, et al. (2008) Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. *Plos One* 3: e3519.
72. Tian C, Kosoy R, Lee A, Ramson M, Belmont JW, et al. (2008) Analysis of East Asia genetic substructure using genome-wide SNP arrays. *Plos One* 3: e3862.
73. Arcos-Burgos M, Muenke M (2002) Genetics of Population Isolates. *Clin Genet* 61: 233–247.
74. Jousilahti P, Vartiainen E, Tuomilehto J, Pekkanen J, Puska P (1998) Role of known risk factors in explaining the difference in the risk of coronary heart disease between eastern and southwestern Finland. *Ann Med* 30: 481–487.
75. Tuomilehto J, Arstila M, Kaarsalo E, Kankaanpää J, Ketonen M, et al. (1992) Acute myocardial infarction (AMI) in Finland-baseline data from the FINMONICA AMI register in 1983–1985. *Eur Heart J* 13: 577–587.
76. Marroni F, Pichler I, De Grandi A, Beu Volpato C, Vogl FD, et al. (2006) Population isolates in South Tyrol and their value for genetic dissection of complex diseases. *Ann Hum Genet* 70: 812–821.
77. Vitart V, Biloglav Z, Hayward C, Janicijevic B, Smolej-Narancic N, et al. (2006) 3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia. *Eur J Hum Genet* 14: 478–487.
78. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls (2007). *Nature* 447: 661–678.
79. Plenge RM, Scielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N Engl J Med* 357: 1199–1209.
80. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, et al. (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 358: 900–909.
81. Gianfrancesco F, Esposito T, Ombra MN, Forabosco P, Maninchedda G, et al. (2003) Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. *Am J Hum Genet* 72: 1479–1491.
82. Prodi DA, Pirastu N, Maninchedda G, Sassu A, Picciau A, et al. (2008) EDA2R is associated with androgenetic alopecia. *J Invest Dermatol* 128: 2268–2270.