

ORIGINAL ARTICLE

# Selective neutrality analysis of 17 STRs in Mediterranean populations

Ignazio Piras<sup>1,2</sup>, Alessandra Falchi<sup>2</sup>, Pedro Moral<sup>3</sup>, Giorgio Paoli<sup>4</sup>, Carla M Calò<sup>2</sup>, Giuseppe Vona<sup>2</sup> and Laurent Varesi<sup>1</sup>

Detection of genes that have been targeted by natural selection is a powerful tool for predicting regions of the genome potentially linked with diseases and of interest in the field of genetic epidemiology. In recent years, several methods to detect patterns of natural selection have been developed. In general, these tests are based on different assumptions and parameters; hence, the detection of outlier loci with more than one statistical approach simultaneously will support the candidate status of a particular locus. In this study, we evaluated the presence of patterns of positive selection in 17 short tandem repeat loci genotyped in six different human populations from the Mediterranean area, for a total of 429 individuals. To identify patterns of selective pressure, we applied three different neutrality tests on the basis of different models, performing pairwise comparisons between populations. Results show the presence of one marker, a (CA)<sub>n</sub> repeat located in exon 29 of the *NOS1* gene, which seems significant in the three different tests in two pairwise comparisons: Sicily vs Morocco and Balearic Islands vs Morocco. This suggests that this locus and its genome localization are candidates for further studies to investigate selective pressure, as well as for association studies.

*Journal of Human Genetics* (2010) 55, 207–214; doi:10.1038/jhgc.2010.7; published online 19 February 2010

**Keywords:**  $F_{ST}$ , LnRH; mediterranean populations; neutrality test; *NOS1*

## INTRODUCTION

Detection of signatures of natural selection in the human genome is a useful tool to identify genes that might underlie variation in disease resistance or drug metabolism. Any form of selection affects some regions of the genome more than others, whereas population history, demography, migration and the mating system will affect the whole genome in the same way.<sup>1</sup> Consequently, genomic regions that are different from the rest of the genome are likely to contain genes that are involved in local adaptation. If selected loci are present in the sample, the variance in single locus  $F_{ST}$  estimates should be higher than if only neutral loci were sampled.

Neutral polymorphic loci are used in making inferences about patterns of differentiation within or among populations of the same or closely related species, because under particular models of population structure, they are related to demographic or historical parameters. Selection may shape the distribution of genetic variation at the so-called 'neutral markers'.<sup>2–5</sup> A neutral locus will respond to selection whenever it is in linkage disequilibrium with other loci that are subject to selection, depending on the local rate of recombination.<sup>6</sup> The process of balancing selection of a locus tends to maintain an elevated level of variation at closely linked neutral loci.<sup>7,8</sup> Other modalities of selection brought a reduction in variability at linked sites. There are two mechanisms responsible: background selection

and genetic hitchhiking.<sup>9,10</sup> Background selection removes deleterious mutations and eliminates variation at linked sites, varying with the recombination rate, the magnitude of selection and the mutation rate.<sup>11,12</sup> Genetic hitchhiking is an outcome of positive selection: if a mutation increases in frequency in a population as a result of selection, it will be accompanied by linked neutral variation.<sup>13</sup>

The first multilocus test to detect selection was developed by Lewontin and Krakauer,<sup>14</sup> on the basis of the sampling distribution of a statistic  $F$  (an estimator of  $F_{ST}$ ). In brief, the variance in  $F_{ST}$  values among loci was used to identify those loci that deviated more than expected. This test was soon shown to be invalid, as the underlying neutral model was far too restrictive because various models of population history have been shown to inflate the expected variance.<sup>15–17</sup> Lewontin and Krakauer's study approach was refined by Bowcock *et al.*<sup>18</sup> and by Beaumont and Nichols.<sup>19</sup> Conditional distribution of  $F_{ST}$  estimates is found to be robust to the details of the neutral model.<sup>19</sup> Moreover, the problem of an unknown population history can be reduced by restricting the analysis to simple scenarios that consider a pair of populations rather than several populations simultaneously.<sup>20</sup>

In recent years, several methods have been developed to detect patterns of natural selection.<sup>21–26</sup> In general, the neutrality test is based on different assumptions and parameters; hence, the detection

<sup>1</sup>Faculty of Science, University of Corsica, Corte, Corsica, France; <sup>2</sup>Department of Experimental Biology, Section Anthropology, University of Cagliari, Monserrato, Italy; <sup>3</sup>Department of Animal Biology, University of Barcelona, Barcelona, Spain and <sup>4</sup>Unit of Anthropology, Department of Ecology and Evolution, University of Pisa, Pisa, Italy  
Correspondence: Dr I Piras, Department of Experimental Biology, Section Anthropology, University of Cagliari, Monserrato, Cagliari 09042, Italy.  
E-mail: is.piras@tiscali.it

Received 5 October 2009; revised 26 December 2009; accepted 23 January 2010; published online 19 February 2010

of outlier loci with more than one statistical approach simultaneously will strengthen the candidate status of a particular locus.<sup>27–29</sup>

As knowledge of markers responding to selection is a useful tool for clinical and molecular anthropology studies, in this study, we evaluate the presence of signatures of natural selection in 17 short tandem repeats genotyped in six human populations from the Mediterranean area, applying three different approaches based on different models. We applied so-called ‘multiple marker based neutrality tests,’ which use information from several loci or genomic regions to construct the neutral null distribution on the basis of the variability characteristics of the markers in the analyzed samples.<sup>30</sup> Following a candidate gene approach, we choose short tandem repeat loci localized in coding and noncoding regions of genes for enzymes of oxidative metabolism, immunity system and erythrocyte membrane components,<sup>31</sup> which are frequent targets of natural selection.<sup>32,33</sup> Erythrocyte membrane components such as spectrin and ankyrin have been associated with spherocytosis, ovalocytosis<sup>34–37</sup> and interactions with different species of plasmodium.<sup>38–40</sup> *TNF* genes encode for proteins belonging to the group of inflammatory cytokines. In particular, they exert an effect in immunostimulation and modulation of host response to infectious agents and cancer cells.<sup>41–42</sup>

Nitric oxide (NO), synthesized by enzymes encoded by *NOS* genes, has an important role in the innate immune response, such as control of viral, bacterial and parasitic infections.<sup>43</sup> It regulates the functioning of the immune system through different mechanisms, dependent and not dependent on cGMP, and modulates cytokine production.<sup>43,44</sup> Genes of oxidative metabolism (*SOD3*, *HO1*, *PON1*) are involved in different processes, including cytoprotective action, removal of radical species and protection against bacterial endotoxins.<sup>45–47</sup>

Microsatellites offer the advantage of a multiallelic marker, which is highly informative. They are unlikely to be targets of natural selection, but linkage to a genomic region that has been the target of selection is expected to cause a deviation from neutral expectations.

## MATERIALS AND METHODS

Samples were taken from 429 individuals from Spain ( $N=126$ ), the Balearic Islands ( $N=62$ ), Sardinia ( $N=90$ ), Tuscany (Italy) ( $N=51$ ), Sicily (Italy) ( $N=47$ ) and Morocco ( $N=53$ ). The samples were from unrelated individuals of both sexes, born and resident in their countries of origin, as their relatives had been for at least three generations. Informed consent was obtained from all participants of the study. The protocols and procedures used in this research were in compliance with the declaration of Helsinki.

DNA was extracted from whole blood using the standard phenol–chloroform technique. DNA samples were amplified using the PCR method and fluorescent primers as previously described.<sup>48–58</sup> The list of polymorphisms analyzed and details of chromosome localization are indicated in Table 1. PCR products were analyzed by an ABI 3730 DNA Analyzer (Applied Biosystems, SA, France). Genotypes were identified using GENESCAN and GENOTYPER softwares (Applied Biosystems).

The Hardy–Weinberg equilibrium was tested using the Markov–Chain Method with 1000000 steps. Standard diversity indexes were calculated according to Nei.<sup>59</sup> To analyze the genetic differentiation between populations, we applied the  $R_{ST}$  pairwise test with 10000 permutations.<sup>60</sup> These analyses were performed using Arlequin 3.1 software.<sup>61</sup>

To detect divergent selection, we applied three methods that identify outlier loci on the basis of various estimators of population divergence,<sup>19,21–23</sup> performing pairwise comparisons between populations to solve the problem of an unknown population history.<sup>20</sup> As all applied neutrality tests are based on different assumptions and parameters, the detection of outlier loci simultaneously with more than one statistical approach will strengthen the candidate status of a particular locus.<sup>27,28</sup>

The first method developed by Beaumont and Nichols<sup>19</sup> and implemented in the software FDIST2, available at <http://www.rubic.rdg.ac.uk/~mab/software.html>,

**Table 1** Chromosome location, percentage of significant *P*-values of Hardy–Weinberg equilibrium and mean number of alleles of markers analyzed

Markers	Chromosome location	Significant <i>P</i> -values (%)	Mean number of alleles
HO-1 (GT)n	22q12-13.1	0.0	6.7
SOD3 (GT)n	4p15.1-15-3	16.7	10.3
CYP11A (TTTTA)n	15q23-24	16.7	5.8
PON1 (GT)n	7q21.3	0.0	2.0
TNFb (GA)n	6p21.3	16.7	8.5
TNFd (GA)n	6p21.3	16.7	20.0
TNFe (GA)n	6p21.3	66.7	3.7
NOS1 (AAT)n in	12q24.2-24.31	16.7	6.0
NOS1 (CA)n 5'	12q24.2-24.31	83.3	14.0
NOS1 (CA)n e29	12q24.2-24.31	66.7	7.5
NOS2 (AAAT)n	17cenq11.2	16.7	3.0
NOS2 (CCTTT)n	17cenq11.2	0.0	4.2
NOS3 (CA)n	7q36	16.7	6.3
ADA (TAAA)n	20q12-13.11	16.7	6.0
ADA (TTTA)n	20q12-13.11	50.0	3.5
SPTA (GT)n	1q21	0.0	9.0
ANK (CA)n	8p11.1-21.1	33.3	4.8

identifies outlying values of  $F_{ST}$  in a plot of  $F_{ST}$  vs heterozygosity using a null distribution based on a symmetrical island model of population structure. This method is based on the assumption that loci showing unusually low or high levels of genetic differentiation are often assured to be subject to natural selection.  $F_{ST}$  is estimated by the statistic  $\beta$ ,<sup>62</sup> whereas the expected  $F_{ST}$  is calculated from the data as the average among loci weighted by their heterozygosity.<sup>63</sup> Coalescent simulations are performed using the island model with 100 islands, and the distribution of  $F_{ST}$  as a function of heterozygosity is characterized by estimating the quantiles of the distribution. We simulated 100000 independent loci, setting sample size to 50 individuals per population in all simulations.

The second method identifies loci that differ in variability from the rest of the genome by calculating the ratio of gene diversity<sup>59</sup> in two populations (RH). LnRH is approximately normally distributed under neutrality.<sup>23</sup> After standardization, 95% of neutral loci are expected to have values between  $-1.96$  and  $1.96$ ; 99% between  $-2.58$  and  $2.58$ ; and 99.9% between  $-3.29$  and  $3.29$ .

The last method is based on a model proposed by Vitalis *et al.*,<sup>21,64</sup> implemented in the software DETSEL 1.0 (<http://www.univ-montp2.fr/~genetix/detsel/detsel.html>). The method is based on a model of population divergence by pure random drift. We performed 100000 coalescent simulations for each pairwise comparison. The nuisance parameters were used in different combinations to generate null distributions with a similar number of allelic states as in the observed data sets. Loci that fall outside the specified ‘probability region,’ compared with the simulated data points, are reported to be potentially affected by selection. This test assumed that no mutations occurred after the divergence of two populations from the common ancestor; hence, we determined whether stepwise-like mutation, in addition to genetic drift, has contributed to genetic differentiation among the studied populations. We calculated whether the observed  $R_{ST}$  value is significantly larger than randomized  $R_{ST}$ ; in this case, the stepwise-like mutations have contributed to the observed differentiation pattern.<sup>65</sup> Analysis was performed with SPAGeDI 1.1,<sup>66</sup> using 20000 permutations.

## RESULTS

Results for the Hardy–Weinberg equilibrium test and the mean number of alleles of markers analyzed are indicated in Table 1. Markers showed departures from Hardy–Weinberg equilibrium in 25.5% of tests. Markers with the highest number of departures were *NOS1* (CA)n 5', with 83.3% of significant *P*-values, followed by *TNFe*

(GA)n and NOS1 (CA)n exon 29 (NOS1 (CA)n e29), with 66.7% of significant departures. On the contrary, HO-1 (GT)n, PON1 (GT)n, NOS2 (CCTTT)n and SPTA (GT)n are in Hardy–Weinberg equilibrium in all populations analyzed (Table 1). The population with the highest number of departures is that from the Balearic Islands (41.2%), whereas Tuscany shows the lowest number (5.8%) (data not shown).

The number of alleles does not show significant heterogeneity across populations (the Kruskal–Wallis test:  $P=0.701$ ). The population with the highest mean number of alleles is Spain ( $8.41 \pm 5.54$ ), whereas the population with the lowest mean number of alleles is Morocco ( $6.29 \pm 4.57$ ).

In the same way, the observed heterozygosity does not show significant heterogeneity across populations (the Kruskal–Wallis test:  $P=0.806$ ).

To evaluate the differences between populations analyzed, we performed an  $R_{ST}$  analysis. We observed a high differentiation across populations. A total of 86.7% of pairwise comparisons are significant, with the exclusion of Sardinia vs Spain and Balearic Islands vs Morocco.

The observed multilocus  $R_{ST}$  values were significantly higher than permuted  $R_{ST}$  values ( $P<0.05$ ) in two pairs of populations: Sardinia–Tuscany and Sicily–Tuscany. For these populations, stepwise-like mutations increased the differentiation between populations.

To apply neutrality tests, we performed pairwise comparisons between the six population samples analyzed, for a total of 15 comparisons for marker, and a total of 255 for the whole set of polymorphisms for each neutrality test (765 comparisons in total). Results for the three analyses are summarized in Table 2. In general, 15.6% of comparisons showed departures from neutrality: 32.8% ( $P<0.05$ ), 37.0% ( $P<0.01$ ) and 30.3% ( $P<0.001$ ). If we consider the different tests, the highest number of comparisons is shown by the F-test (34.5%), whereas the  $F_{ST}$  test and the LnRH test showed 6.7 and 5.9% significant comparisons, respectively. Comparisons between the

three analyses are shown in Table 3. Markers that show departures from neutrality in two tests are SOD3 (GT)n (53.3%); TNFe (GA)n (40.0%); NOS1 (CA)n e29 (20.0%); NOS2 (AAAT)n and NOS2 (CCTTT)n (13.3%); and TNFb (GA)n (6.6%). Markers that show departures from neutrality in three tests are TNFe (GA)n (Sicily vs Tuscany); NOS1 (CA)n e29 (CA) (Sicily vs Morocco); and (Balearic Islands vs Morocco). Results of the analysis of marker NOS1 (CA)n e29 are shown in Figures 1–3. Otherwise, we can detect populations under selection, if the population is significant in all pairwise comparisons. For marker TNFe (GA)n, the population from Tuscany is significant in three tests in one comparison (with the Sicily population), but significant in two tests in three comparisons (with Sardinia, Morocco and Balearic Islands). For other loci identified, the Moroccan population is significant in the three tests in two pairwise comparisons, and in two tests in other comparisons.

## DISCUSSION

In this study, we evaluated the presence of patterns of divergent selection using a set of 17 polymorphic microsatellite markers located in different chromosomes. We genotyped a total of 429 individuals from six different human Mediterranean populations. We used three methods based on different models, considering markers that showed significant results in all neutrality tests used, according to an approach adopted by other authors.<sup>27,28</sup>

When we compared the results obtained using the three methods, the markers responding to selection were TNFe (GA)n in comparison between Sicily and Tuscany, and NOS1 (CA)n e29 in comparison between Sicily and Morocco and Balearic Islands and Morocco. If we consider the number of significant comparisons of such populations, we can underlie that the Moroccan population, for NOS1 (CA)n e29, is significant in all comparisons with other populations, in particular in two cases with the three tests, and in other cases in two tests. The same consideration can be made for Tuscany for the marker TNFe.

It is necessary to reject results for Sicily–Tuscany, because the observed  $R_{ST}$  value is significantly higher than randomized  $R_{ST}$  ( $P<0.05$ ), indicating that stepwise-like mutation created a difference between these two populations. Consequently, the results obtained with the F-test, which assumes divergence by pure genetic drift, could be false positives.

We conclude that the only marker that responds to natural selection is NOS1 (CA)n e29, in population comparisons between Morocco and Balearic Islands and Sicily. In particular, the population that responds to selection seems to be the Moroccan population because its comparisons are always significant in two or three tests with other populations.

A limitation of this study is the number of markers used with respect to analysis performed with microarray data. In fact, for this study, we used a candidate gene approach, and hence the study suffers from two limitations of such a type of analysis: *a priori* hypothesis on the candidate gene used and the confounding effect of the genetic drift.<sup>67</sup>

Despite these limitations, the results we obtained are validated by the method used, because the simultaneous detection of outlier loci in different tests based on different assumptions and parameters reduces the possibility of false positives.<sup>27–29</sup>

Moreover, we highlight that results obtained with regard to signals of natural selection in the genome are similar to those of some previously published papers in which genomic scans with microsatellite markers have been performed.<sup>27,68,69</sup> In these papers, the percentage of markers responding to selection ranges from 2.1% to 10.6%, whereas we detected a value of 5.9%.

**Table 2** Number of significant comparisons for each neutrality test applied

	$F_{ST}$ test		LnRH		F-test	
	N	%	N	%	N	%
HO-1 (GT)n	0	0.0	0.0	0.0	31.0	68.9
SOD3 (GT)n	14	31.1	0.0	0.0	25.0	55.6
CYP11A (TTTTA)n	0	0.0	0.0	0.0	2.0	4.4
PON1 (GT)n	0	0.0	0.0	0.0	13.0	28.9
TNFb (GA)n	2	4.4	0.0	0.0	4.0	8.9
TNFd	0	0.0	1.0	2.2	9.0	20.0
TNFe	1	2.2	25.0	55.6	19.0	42.2
NOS1 (AAT)n in	0	0.0	0.0	0.0	19.0	42.2
NOS1 (CA)n 5'	4	8.9	0.0	0.0	14.0	31.1
NOS1 (CA)n e29	2	4.4	14.0	31.1	26.0	57.8
NOS2 (AAAT)n	0	0.0	14.0	31.1	27.0	60.0
NOS2 (CCTTT)n	3	6.7	0.0	0.0	15.0	33.3
NOS3 (CA)n	9	20.0	0.0	0.0	3.0	6.7
ADA (TAAA)n	1	2.2	0.0	0.0	8.0	17.8
ADA (TTTA)n	0	0.0	8.0	17.8	14.0	31.1
SPTA (GT)n	0	0.0	0.0	0.0	15.0	33.3
ANK (CA)n	0	0.0	0.0	0.0	2.0	4.4
TOT	36	5.0	62	8.1	246	32.2

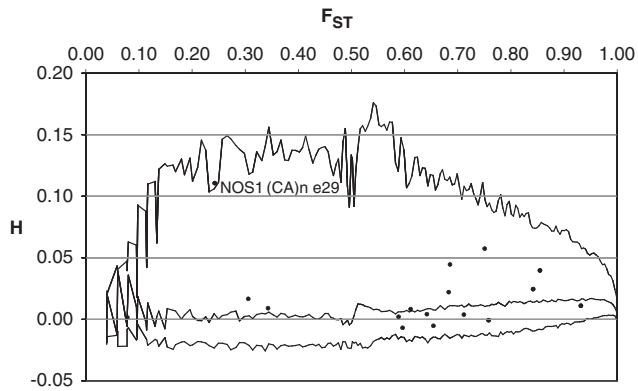
**Table 3 Results of each pairwise comparison of neutrality tests applied**

	HO-1	SOD3	CYP11A	PON1	TNFb	TNFd	TNFe	NOS1in	NOS15'	NOS1e29	NOS2d	NOS2p	NOS3	ADA1	ADA2	SPTA1	ANK1
SA-SI																	
a		***			*							*					
b	**	***			**		***					**				**	
c																	
SA-TU																	
a																	
b	**				*		**										
c							***										
SA-BI																	
a		***						*									
b	**	***						**			*			*			
c											**						
SA-SP																	
a		*										*					
b	***	***			*						*	***					
c											***						
SA-MO																	
a		*															
b	**	**			*		*	**	*	***							
c							*			***							
SI-TU																	
a		***					*										
b		***					***									**	
c							***										
SI-BI																	
a							*		***		*		**				
b							***										
c							***				***						
SI-SP																	
a							***			*							
b							***			***					*		***
c		**					***										
SI-MO																	
a																	
b	***									*	*						*
c										***	***						***

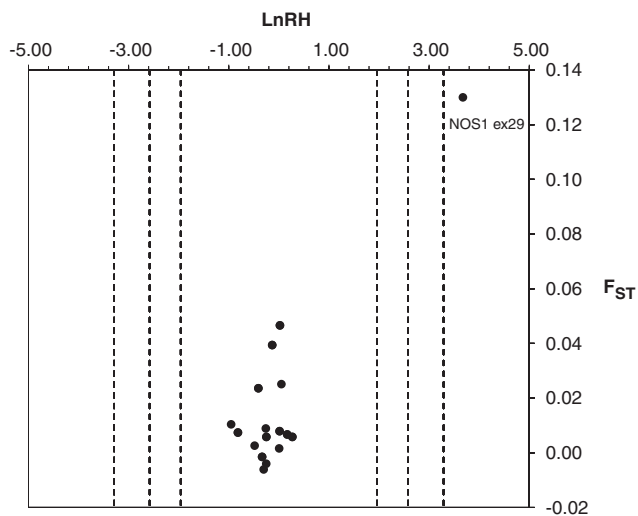
**Table 3 Continued**

	HO-1	SOD3	CYP11A	PON1	TNFb	TNFd	TNFe	NOS1in	NOS15'	NOS1e29	NOS2d	NOS2p	NOS3	ADA1	ADA2	SPTA1	ANK1
<i>TU-BI</i>																	
a		***					**	*	**		*			**			**
b		***		*			**										
c							**										
<i>TU-SP</i>												**					
a		*					**				**						
b	**	**		**			**				**	**					
c																	
<i>TU-MO</i>													*				
a		**					***	**	***								
b	***	***		**		*	***	**	***								
c							**		**								
<i>BI-SP</i>																	
a																	
b		**		*			**	**	*	*	*		*	**		**	
c																	
<i>BI-MO</i>																	
a									*	*							
b	***			**		*		*	***	***	*						
c									**	**							
<i>SP-MO</i>																	
a																	**
b	**	**		**			***	**	***	***	*		**				**
c									**	**							**

Abbreviations: a, F<sub>ST</sub> test; BI, Balearic Islands; b, F-test; c, LnRH test; MO, Morocco; SA, Sardinia; S, Sicily; SP, Spain; TU, Tuscany.  
\**P*<0.05; \*\**P*<0.01; \*\*\**P*<0.001.



**Figure 1** Results of the  $F_{ST}$  test for Sicily and Morocco for all markers. Each dot corresponds to a marker. The two external lines correspond to 95% confidence intervals, and the internal line corresponds to median values. Markers out of area (in this case, *NOS1* (CA)n e29) are significant at  $P < 0.05$ .



**Figure 2** Results of the LnRH test between Sicilian and Moroccan populations. Dotted lines indicate 95.0%, 99.0% and 99.9% confidence intervals.

The *NOS1* gene codifies for neuronal nitric oxide synthase (NOS), which synthesizes for NO responsible for neurotransmission of NANC synapsis; moreover, it is implied in the physiopathology of several cerebral damages. In addition to the neurons of the central and peripheral nervous system, *NOS1* is expressed in skeletal muscles and in the cells of the epithelium of the respiratory system.<sup>70,71</sup> A recent study has reported the first example of *NOS1* being involved in the elimination of an infection in mice.<sup>72</sup> The *NOS1* (CA)n repeat localized in the 3' untranslated region of exon 29 and the existence of allelic mRNA sequence variation have been reported.<sup>73</sup> An association has also been demonstrated between asthma and this dinucleotide repeat polymorphism<sup>74,75</sup> confirming a functional role of this short tandem repeat. NO has been shown to have a crucial role in immunoregulation and it is implicated in host nonspecific defense in a variety of infections such as malaria, toxoplasmosis, leishmaniosis, trypanosomosis and schistosomosis.<sup>76</sup>

Recently, we suggested a selective pressure by endemic malaria exerted on *NOS1* (CA)n e29 alleles that correlated to NO overproduction in Corsica.<sup>77</sup> In this paper, we showed that alleles previously correlated with NO overproduction (alleles 16 and 17) are more represented in a sample of  $\beta^{039}$  carriers than in control samples, suggesting a selective effect of malaria on alleles that correlated with NO overproduction. If we consider allele frequencies in populations analyzed in the present study, we observe that Morocco has the highest frequencies of alleles 16 and 17 with respect to other populations, suggesting a positive selection for these alleles. This result is according to the hypothesis of malaria infection,<sup>77</sup> because of a high impact of endemic malaria in the North African region compared with the South European region.<sup>78</sup> In Sardinia, where malaria had a high prevalence since eradication immediately after World War II, signals of natural selection in the *NOS1* gene were absent, probably due to genetic drift and isolation, which could have modified allele frequencies, and also due to the phenomena of local adaptation.

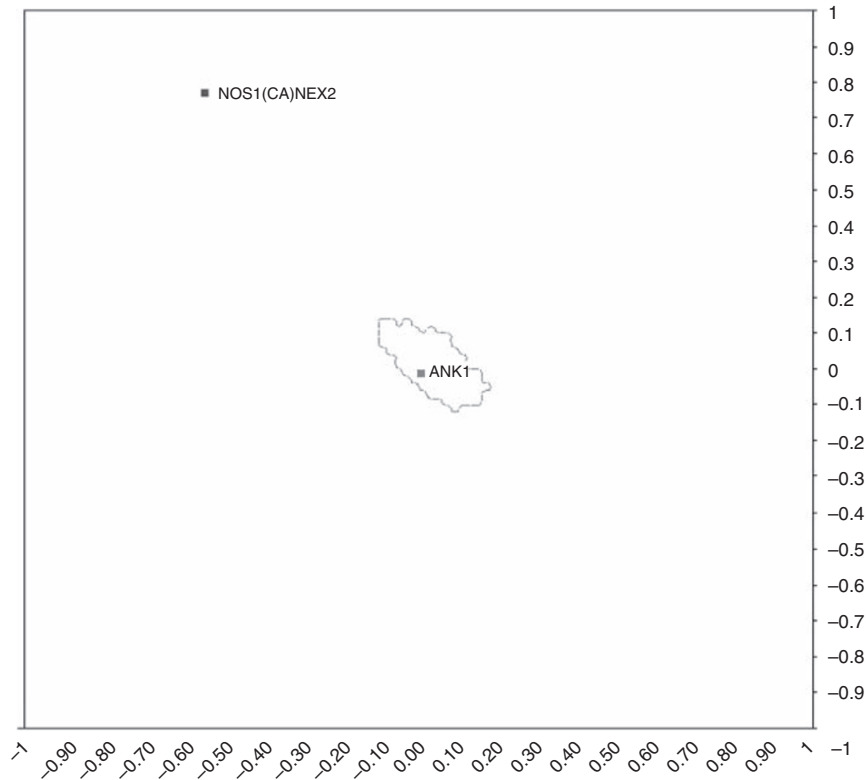
In recent years, the development of genome-wide technologies has permitted the analysis of natural selection patterns with high resolution. To date, 21 genome-wide scans for positive selection have been performed in human populations, producing maps of positive selection. Several analyses have been performed in worldwide populations, and several regions have been detected as targets of positive selection.<sup>67</sup> Positive selection at the *NOS1* gene, located in chr12 at 117650979–117799582, has not been detected in these studies. This could be because of local adaptation phenomena, because, as observed in previous studies, signatures of natural selection are not uniformly distributed across populations, but rather show clear spatial heterogeneity. We use different and more restricted populations than other studies that have used HapMap samples or samples from large geographical areas. There are a lot of papers describing spatially varying patterns of selection.<sup>79</sup> Local adaptation is due to the environmental heterogeneity that the human population has been confronted with throughout the world after the 'out of Africa' migration. Moreover, previous analyses used different statistical methods, which is an important factor in interpreting results across studies,<sup>67</sup> and it is important to emphasize that the lack of overlapping of regions under selection detected in different studies is quite common. In 21 studies analyzed by Akey *et al.*,<sup>67</sup> only 722 regions over 5110 totally identified were detected in two or more different studies; 271 were identified in three or more studies and 129 regions were identified in four or more studies.

Considering the performance of three neutrality tests, the F-test performed reveals the highest number of significant comparisons (34.5 vs 6.7% and 5.9% of the  $F_{st}$  test and LnRH, respectively). Similar results were obtained by Vasemagi *et al.*,<sup>28</sup> considering a screening of 17 genomic and 78 expressed sequence tag-associated mini and microsatellites. In their work, discrepancy is lower if they consider a small spatial scale rather than a large spatial scale, explaining this with the effect of mutations and by the moderate effect of genetic drift on the genetic parameters used to infer the outlier loci.<sup>19,21,22</sup> In our study, inconsistency of results in this way can be explained for only two pairs of populations (Sardinia vs Tuscany; Sicily vs Tuscany), as indicated by the  $R_{ST}$  permutation test of Hardy *et al.*<sup>65</sup>

An assumption that has possibly been violated is the absence of migration. The F-test assumed that no migrants have been exchanged after the divergence of two populations, even if moderate levels of migration do not increase the false positive results.<sup>21</sup>

In conclusion, the marker that we could consider under divergent selection is *NOS1* (CA)n e29. We should underline that significant deviation from neutral expectations using one or multiple tests does





**Figure 3** Results of the F test between Sicilian and Moroccan populations with regard to ANK1 (CA)<sub>n</sub> and NOS1 (CA)<sub>n</sub> e29. Markers out of area show a level of significance  $P < 0.01$ . F1 and F2 values are represented in X and Y axes.

not necessarily mean that a particular locus has been affected by selection. These significant results only raise the candidate status of a particular locus and does not demonstrate selection *per se*,<sup>22,80,81</sup> because the violation of test assumptions is another factor potentially producing false positive results. This candidate locus will serve as a basis for further sequence analysis to validate the role of divergent selection in this gene.

Several authors have argued that positive selection might be frequent in the genomes of humans and other organisms.<sup>82</sup> If this is true, we have the necessary statistical methods for identifying *loci* that have undergone selection on the basis of comparative data. It will be possible to make systematic searches for genes that have undergone positive selection in the lineage leading to humans and identify the adaptive changes at the molecular level that were important in the evolution of modern humans.

Identifying selection in the genome might very well become one of our most powerful tools for identifying causes for species-specific differences and for identifying genomic regions of functional, and perhaps, medical importance.

#### ACKNOWLEDGEMENTS

This research was supported by European program Interreg IIIA (GV and LV). We thank two anonymous reviewers for constructive comments on the paper.

- 1 Cavalli-Sforza, L. L. Population structure and human evolution. *Proc. R. Soc. Lond. B. Biol. Sci.* **164**, 362–379 (1966).
- 2 Singh, R. S. & Rhomberg, L. R. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics* **117**, 255–271 (1987).

- 3 Choudhary, M., Coulthart, M. B. & Singh, R. S. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VI. Patterns and processes of genic divergence between *D. melanogaster* and its sibling species, *Drosophila simulans*. *Genetics* **130**, 843–853 (1992).
- 4 Wilding, C. S., Grahame, J. & Mill, P. J. Nuclear DNA restriction site polymorphisms and the phylogeny and population structure of an intertidal snail species complex (*Littorina*). *Hereditas* **133**, 9–18 (2000).
- 5 Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
- 6 Raymond, M., Vaanto, R. L., Thomas, F., Rousset, R., de Meeüs, T. & Renaud, F. Heterozygote deficiency in the mussel *Mytilus edulis* species complex revisited. *Mar. Ecol. Prog. Ser.* **156**, 225–237 (1997).
- 7 Strobeck, C. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**, 545–555 (1983).
- 8 Hudson, R. R. & Kaplan, N. L. The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840 (1988).
- 9 Eriksson, A., Fernström, P., Mehlig, B. & Sagitov, S. An accurate model for genetic hitchhiking. *Genetics* **178**, 439–451 (2008).
- 10 McCollum, A. M., Basco, L. K., Tahar, R., Udhayakumar, V. & Escalante, A. A. Hitchhiking and selective sweeps of *Plasmodium falciparum* sulfadoxine and pyrimethamine resistance alleles in a population from central Africa. *Antimicrob. Agents Chemother.* **52**, 4089–4097 (2008).
- 11 Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
- 12 Hudson, R. R. & Kaplan, N. L. Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).
- 13 Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
- 14 Lewontin, R. C. & Krakauer, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195 (1973).
- 15 Nei, M. & Chakravarti, A. Drift variances of  $F_{ST}$  and  $G_{ST}$  statistics obtained from a finite number of isolated populations. *Theor. Popul. Biol.* **11**, 307–325 (1977).
- 16 Nei, M. & Maruyama, T. Lewontin–Krakauer test for neutral genes. *Genetics* **80**, 395 (1975).
- 17 Robertson, A. Remarks on the Lewontin–Krakauer test. *Genetics* **80**, 396 (1975).
- 18 Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. et al. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl Acad. Sci. USA* **88**, 839–843 (1991).

- 19 Beaumont, M. A. & Nichols, R. A. Evaluating loci for use in the genetic analysis of population structure. *Proc. Royal. Soc. Lon. B* **263**, 1619–1626 (1996).
- 20 Tsakas, S. & Krimbas, C. B. Testing the heterogeneity of F values: a suggestion and a correction. *Genetics* **84**, 399–401 (1976).
- 21 Vitalis, R., Dawson, K. & Boursot, P. Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811–1823 (2001).
- 22 Schlotterer, C. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753–763 (2002).
- 23 Kauer, M. O., Dieringer, D. & Schlötterer, C. A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics* **165**, 1137–1148 (2003).
- 24 Beaumont, M. A. & Balding, D. J. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**, 969–980 (2004).
- 25 Drummond, A. J. & Suchard, M. A. Fully Bayesian tests of neutrality using genealogical summary statistics. *BMC Genet.* **9**, 68 (2008).
- 26 Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *Plos Biol.* **5**, e171 (2007).
- 27 Storz, J. F., Payseur, B. A. & Nachman, M. W. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**, 1800–1811 (2004).
- 28 Vasemagi, A., Nilsson, J. & Primmer, C. R. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. Evol.* **22**, 1067–1076 (2005).
- 29 Zhai, W., Nielsen, R. & Slatkin, M. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol. Biol. Evol.* **26**, 273–283 (2009).
- 30 Vasemagi, A. & Primmer, C. R. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* **14**, 3623–3642 (2005).
- 31 Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G. P., Menozzi, G. *et al.* Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res.* **19**, 199–212 (2009).
- 32 Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
- 33 Bamshad, M. & Wooding, S. P. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4**, 99–111 (2003).
- 34 Gallagher, P. G., Sabatino, D. E., Basserres, D. S., Nilson, D. M., Wong, C., Cline, A. P. *et al.* Erythrocyte ankyrin promoter mutations associated with recessive hereditary spherocytosis cause significant abnormalities in ankyrin expression. *J. Biol. Chem.* **276**, 41683–41689 (2001).
- 35 Delaunay, J. Molecular basis of red cell membrane disorders. *Acta Haematol.* **108**, 210–218 (2002).
- 36 Eber, S. W., Gonzalez, J. M., Lux, M. L., Scarpa, A. L., Tse, W. T., Dornwell, M. *et al.* Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis. *Nat. Genet.* **13**, 214–218 (1996).
- 37 Ozcan, R., Jarolim, P., Lux, S. E., Ungewickell, E. & Eber, S. W. Simultaneous (AC)<sub>n</sub> microsatellite polymorphism analysis and single-stranded conformation polymorphism screening is an efficient strategy for detecting ankyrin-1 mutations in dominant hereditary spherocytosis. *Br. J. Haematol.* **122**, 669–677 (2003).
- 38 Schulman, S., Roth, E.F.Jr., Cheng, B., Rybicki, A. C., Sussman, I. I., Wong, M. *et al.* Growth of *Plasmodium falciparum* in human erythrocytes containing abnormal membrane proteins. *Proc. Natl Acad. Sci. USA* **87**, 7339–7343 (1990).
- 39 Dua, M., Raphael, P., Sijwali, P. S., Rosenthal, P. J. & Hanspal, M. Recombinant falcipain-2 cleaves erythrocyte membrane ankyrin and protein 4.1. *Mol. Biochem. Parasitol.* **116**, 95–99 (2001).
- 40 Hanspal, M., Dua, M., Takakuwa, Y., Chishti, A. H. & Mizuno, A. *Plasmodium falciparum* cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins at late stages of parasite development. *Blood* **100**, 1048–1054 (2002).
- 41 Old, L. J. Tumor necrosis factor (TNF). *Science* **230**, 630–632 (1985).
- 42 Havell, E. A. Production of tumor necrosis factor during murine listeriosis. *J. Immunol.* **139**, 4225–4231 (1987).
- 43 Bogdan, C. Nitric oxide and the immune response. *Nat. Immunol.* **2**, 907–916 (2001).
- 44 van der Veen, R. C. Nitric oxide and T helper cell immunity. *Int. Immunopharmacol.* **1**, 1491–1500 (2001).
- 45 Applegate, L. A., Luscher, P. & Tyrrell, R. M. Induction of heme oxygenase: a general response to oxidant stress in cultured mammalian cells. *Cancer Res.* **51**, 974–978 (1991).
- 46 Johnson, K. J., Ward, P. A., Gorainick, S. & Osborn, M. J. Isolation from human serum of an inactivator of bacterial lipopolysaccharide. *Am. J. Pathol.* **88**, 559–574 (1977).
- 47 Stocker, R., Yamamoto, Y., McDonagh, A. F., Glazer, A. N. & Ames, B. N. Bilirubin is an antioxidant of possible physiological importance. *Science* **235**, 1043–1046 (1987).
- 48 Mohandas, T., Sparkes, R. S., Suh, E. J. & Hershfield, M. S. Regional localization of the human genes for S-adenosylhomocysteine hydrolase (cen—q131) and adenosine deaminase (q131—qter) on chromosome 20. *Hum. Genet.* **66**, 292–295 (1984).
- 49 Hentati, A., Hu, P., Asgharzadeh, S. & Siddique, T. Dinucleotide repeat polymorphism at the human erythroid  $\alpha$  spectrin (*SPTA1*) locus. *Hum. Mol. Genet.* **1**, 218 (1992).
- 50 Polymeropoulos, M. H., Rath, D. S., Xiao, H. & Merrill, C. R. Dinucleotide repeat polymorphism at the human ankyrin gene (*ANK1*). *Nucleic Acids Res.* **19**, 969 (1991).
- 51 Udalova, I. A., Nedospasov, S. A., Webb, G. C., Chaplin, D. D. & Turetskaya, R. L. Highly informative typing of the human *TNF* locus using six adjacent polymorphic markers. *Genomics* **16**, 180–186 (1993).
- 52 Clendenning, J. B., Humbert, R., Green, E. D., Wood, C., Traver, D. & Furlong, C. E. Structural organization of the human *PON1* gene. *Genomics* **35**, 586–589 (1996).
- 53 Glenn, C. L., Wang, W. Y. & Morris, B. J. Different frequencies of inducible nitric oxide synthase genotypes in older hypertensives. *Hypertension* **33**, 927–932 (1999).
- 54 Levesque, M. C., Hobbs, M. R., Anstey, N. M., Vaughn, T. N., Chancellor, J. A., Pole, A. *et al.* Nitric oxide synthase type 2 promoter polymorphisms, nitric oxide production, and disease severity in Tanzanian children with malaria. *J. Infect. Dis.* **180**, 1994–2002 (1999).
- 55 Lembo, G., De Luca, N., Battagli, C., Iovino, G., Aretini, A., Musicco, M. *et al.* A common variant of endothelial nitric oxide synthase (Glu298Asp) is an independent risk factor for carotid atherosclerosis. *Stroke* **32**, 735–740 (2001).
- 56 San Millan, J. L., Sancho, J., Calvo, R. M. & Escobar-Morreale, H. F. Role of the pentanucleotide (TTTA)<sub>n</sub> polymorphism in the promoter of the *CYP11A* gene in the pathogenesis of hirsutism. *Fertil. Steril.* **75**, 797–802 (2001).
- 57 Kaneda, H., Ohno, M., Taguchi, J., Togo, M., Hashimoto, H., Ogasawara, K. *et al.* Heme oxygenase-1 gene promoter polymorphism is associated with coronary artery disease in Japanese patients with coronary risk factors. *Arterioscler. Thromb. Biol.* **22**, 1680–1685 (2002).
- 58 Stern, L. F., Chapman, N. H., Wijsman, E. M., Altherr, M. R. & Rosen, D. R. Assignment of *SOD3* to human chromosome band 4p15.3→p15.1 with somatic cell and radiation hybrid mapping, linkage mapping, and fluorescent *in-situ* hybridization. *Cytogenet. Genome Res.* **101**, 178 (2003).
- 59 Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590 (1978).
- 60 Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
- 61 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinf. Online* **1**, 47–50 (2005).
- 62 Cockerham, C. C. & Weir, B. S. Estimation of gene flow from F-statistics. *Evolution* **47**, 855–863 (1993).
- 63 Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
- 64 Vitalis, R., Dawson, K., Boursot, P. & Belkhir, K. DetSel 1.0: a computer program to detect markers responding to selection. *J. Hered.* **94**, 429–431 (2003).
- 65 Hardy, O. J., Charbonnel, N., Freville, H. & Heuertz, M. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**, 1467–1482 (2003).
- 66 Hardy, O. J. & Vekemans, X. SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**, 618–620 (2002).
- 67 Akey, J. M. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* **19**, 711–722 (2009).
- 68 Kayser, M., Brauer, S. & Stoneking, M. A Genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* **20**, 893–900 (2003).
- 69 Payseur, B. A., Cutter, A. D. & Nachman, M. W. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. *Mol. Biol. Evol.* **19**, 1143–1153 (2002).
- 70 Asano, K., Chee, C. B., Gaston, B., Lilly, C. M., Gerard, C., Drazen, J. M. *et al.* Constitutive and inducible nitric oxide synthase gene expression, regulation, and activity in human lung epithelial cells. *Proc. Natl Acad. Sci. USA* **91**, 10089–10093 (1994).
- 71 Forstermann, U., Boissel, J. P. & Kleinert, H. Expressional control of the 'constitutive' isoforms of nitric oxide synthase (*NOS I* and *NOS III*). *FASEB J.* **12**, 773–790 (1998).
- 72 Li, E., Zhou, P. & Singer, S. M. Neuronal nitric oxide synthase is necessary for elimination of *Giardia lamblia* infections in mice. *J. Immunol.* **176**, 516–521 (2006).
- 73 Hall, A. V., Antoniou, H., Wang, Y., Cheung, A. H., Arbus, A. M., Olson, S. L. *et al.* Structural organization of the human neuronal nitric oxide synthase gene (*NOS1*). *J. Biol. Chem.* **269**, 33082–33090 (1994).
- 74 Grasmann, H., Yandava, C. N., Storm van's Gravesande, K., Deykin, A., Pillari, A., Ma, J. *et al.* A neuronal NO synthase (*NOS1*) gene polymorphism is associated with asthma. *Biochem. Biophys. Res. Commun.* **272**, 391–394 (2000).
- 75 Martínez, B., Barrios, K., Vergara, C., Mercado, D., Jiménez, S., Gusmão, L. *et al.* A *NOS1* gene polymorphism associated with asthma and specific immunoglobulin E response to mite allergens in a Colombian population. *Int. Arch. Allergy Immunol.* **144**, 105–113 (2007).
- 76 Rivero, A. Nitric oxide: an antiparasitic molecule of invertebrates. *Trends Parasitol.* **22**, 219–225 (2006).
- 77 Piras, I., Falchi, A., Melis, A., De Cian, M. C., Calò, C. M., Vona, G. *et al.* High frequencies of short alleles of *NOS1* (CA)<sub>n</sub> polymorphism in  $\beta$ 29 carriers from Corsica Island (France). *Exp. Mol. Pathol.* **86**, 136–137 (2009).
- 78 Hay, S. I., Guerra, A. C., Tatern, A. J., Noor, A. M. & Snow, R. W. The global distribution and population at risk of malaria: past, present, and future. *The Lancet Infect. Dis.* **4**, 327–336 (2004).
- 79 Ronald, J. & Akey, J. M. Genome-wide scans for loci under selection in humans. *Hum. Genomics* **2**, 113–125 (2005).
- 80 Vigouroux, Y., McMullen, M., Hittinger, C. T., Houchins, K., Schulz, L., Kresovich, S. *et al.* Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl Acad. Sci. USA* **99**, 9650–9655 (2002).
- 81 Campbell, D. & Bernatchez, L. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* **21**, 945–956 (2004).
- 82 Kreitman, M. & Akashi, H. Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.* **26**, 403–422 (1995).