

UNIVERSITA' DEGLI STUDI DI CAGLIARI

FACOLTA' DI MEDICINA E CHIRURGIA

DIPARTIMENTO DI SCIENZE BIOMEDICHE E
BIOTECNOLOGIE

Dottorato di Ricerca in:

“Terapia Pediatrica e Farmacologia dello Sviluppo”

**VARIABILITA' GENETICA DEL CROMOSOMA Y IN
SARDEGNA:
IMPLICAZIONI NEGLI STUDI DI ASSOCIAZIONE
SULLE MALATTIE MULTIFATTORIALI**

Tutore:
Prof. Francesco Cucca

Autore:
D.ssa Daniela Contu

Anno Accademico 2006-2007

INDICE

INTRODUZIONE	<i>Pag. 3</i>
MATERIALI E METODI	<i>Pag. 16</i>
Selezione dei campioni	<i>Pag. 16</i>
Genotipizzazione	<i>Pag. 20</i>
SNPs	<i>Pag. 20</i>
STRs	<i>Pag. 24</i>
Analisi Statistica	<i>Pag. 26</i>
RISULTATI	<i>Pag. 29</i>
DISCUSSIONE	<i>Pag. 44</i>
BIBLIOGRAFIA	<i>Pag. 57</i>

**VARIABILITA' GENETICA DEL CROMOSOMA Y IN
SARDEGNA:
IMPLICAZIONI NEGLI STUDI DI ASSOCIAZIONE
SULLE MALATTIE MULTIFATTORIALI**

INTRODUZIONE

Da sempre, la popolazione sarda ha suscitato l'interesse dei genetisti umani. La motivazione di un tale interesse va ricercata in alcune caratteristiche di questa popolazione quali, ad esempio, la peculiare distribuzione di alleli a svariati loci o le numerose malattie su base genetica particolarmente frequenti nell'Isola, peculiarità queste, legate verosimilmente proprio all'insularità. In particolare, tali caratteristiche hanno offerto, ad esempio, l'opportunità di studiare l'impatto della selezione naturale nel determinare resistenza alla malaria [1] e di chiarire aspetti rilevanti delle basi molecolari di malattie monogeniche

comuni in Sardegna, quali la Talassemia [2], il Morbo di Wilson [3] e l'APECED [4]. Un ulteriore motivo d'interesse è rappresentato dall'elevata incidenza, in questa popolazione, di malattie multifattoriali autoimmuni quali il diabete di tipo 1 e la sclerosi multipla. La Sardegna, in particolare, condivide con la Finlandia la più alta incidenza di diabete di tipo 1 al mondo [5]. Ciò rappresenta una rimarchevole eccezione al gradiente nord-sud mostrato dall'incidenza di tale patologia. Anche la sclerosi multipla, presenta nei sardi un'incidenza considerevolmente più elevata rispetto alle altre popolazioni europee [6]. A parte le similitudini sul piano epidemiologico, il diabete di tipo 1 e la sclerosi multipla mostrano in Sardegna una tendenza alla co-morbidity negli stessi individui e alla co-occorrenza nelle stesse famiglie [7].

Ciò suggerisce che in Sardegna siano frequenti alleli di suscettibilità nei confronti di entrambe le patologie. Dunque, la popolazione sarda si configura come particolarmente adatta per la ricerca di tali alleli attraverso degli studi di associazione estesi a tutto il genoma (Genome Wide Associations-GWA Studies). Ciò risulta particolarmente rilevante se si tiene conto del fatto che proprio il recente avvento di tali studi, basati sull'uso di centinaia di migliaia di marcatori polimorfici, ha permesso di incrementare sensibilmente il potere di identificare le varianti di rischio per le malattie multifattoriali.

Tuttavia, la conoscenza dei parametri chiave della struttura genetica e della passata storia demografica di questa popolazione ancora parziale. Quanto è antica la popolazione sarda? Qual'è la spiegazione per i numerosi effetti fondatori osservati nei diversi studi genetici

effettuati finora? Qual'è l'impatto delle invasioni ad opera di popoli "stranieri" sulla struttura genetica e sull'eventuale sub-struttura della popolazione sarda attuale? E, conseguentemente, esiste o no sub-struttura ed, eventualmente, di quale grado, in questa popolazione? La risposta a tali domande risulta essenziale al fine di predisporre un disegno sperimentale ottimale per gli studi di associazione caso-controllo che, per ragioni economiche, rappresentano un disegno sperimentale pressoché obbligato in studi GWA miranti all'identificazione delle varianti genetiche che sottostanno a malattie multifattoriali. Infatti, i test d'associazione basati sull'analisi delle famiglie non sono influenzati dalla presenza di sub-struttura di popolazione, ma implicano la genotipizzazione degli affetti e dei genitori, quindi un terzo di genotipi in più rispetto allo studio caso/controllo.

Viceversa, l'impatto della struttura genetica di una popolazione negli studi di associazione di tipo caso controllo può essere duplice: si possono ottenere associazioni spurie di alleli con fenotipi di malattia oppure si possono sottovalutare delle associazioni vere. Ciò si verifica quando le frequenze alleliche differiscono in modo così significativo tra sub-popolazioni da non essere equamente rappresentate tra casi e controlli. La gran parte degli studi con questo disegno sperimentale cerca di ovviare a tale problema attraverso un campionamento dei casi e dei controlli eseguito su gruppi di individui che condividono la stessa nazionalità o la stessa appartenenza ad una data etnia. Tuttavia, alcuni studi dimostrano che ciò non è sufficiente a proteggersi dall'insidia rappresentata dall'esistenza di sub-struttura. Un caso assai interessante è stato riportato da Helgason e colleghi, i quali hanno dimostrato che anche in una popolazione

ritenuta relativamente omogenea e che si caratterizza per essere un isolato genetico, quale è la popolazione islandese, esiste una considerevole suddivisione regionale, verosimilmente legata alle colonizzazioni ad opera di popolazioni distinte, provenienti dalla Scandinavia e dalle Isole Britanniche, verificatesi circa 1.100 anni fa [8]. Gli effetti confondenti per gli studi di associazione legati alla stratificazione di popolazione possono essere corretti con procedure statistiche, per esempio utilizzando controlli genomici, il che implica però perdita di potere statistico.

Ciò suggerisce con forza che la conoscenza dell'entità e del pattern di differenziazione genetica regionale, in determinate popolazioni, possa, non solo, minimizzare il rischio di ottenere risultati falsi positivi negli studi di associazione, ma anche consentire di predisporre il disegno sperimentale ottimale per quella popolazione.

Tornando alla popolazione sarda, per quanto attiene alla sua “età”, esistono evidenze archeologiche che testimoniano un’occupazione pre-neolitica, dopo la quale, durante il periodo Neolitico (8.000 anni fa), la popolazione ha subito un graduale incremento di taglia. Durante l’età del Bronzo (3.600-2.200 anni fa) si sviluppò un’avanzata civiltà, la cui caratteristica più saliente è rappresentata dalla costruzione dei Nuraghe. Durante questo periodo si stima che la popolazione fosse equivalente a circa 300.000 unità [9] ed che sia rimasta relativamente costante durante le successive dominazioni Fenicio-Punica (535 - 238 a.C.) e Romana (238 a.C. - 476 d.C.). Le dimensioni della popolazione sarda non hanno subito un incremento significativo fino a circa 300 anni fa. Infatti, nel 1627, un censimento, effettuato per motivi fiscali, stimò la popolazione come equivalente a 297.000 unità. Nel 1781 la Sardegna contava 431.000 abitanti, mentre nel 1871 i

sardi erano 636.000. I sardi divennero 1.000.000 nel 1936 ed oggi l'Isola conta 1.632.000 persone.

Benché interessanti, questi dati archeologici e storici non forniscono un quadro completo della storia demografica della popolazione sarda e dell'impatto che tali eventi hanno avuto sulla struttura della popolazione attualmente esistente. Inoltre, essi non aiutano a capire gli eventi più antichi che hanno caratterizzato la fondazione di questa popolazione.

La genetica, per la sua stessa natura, può, laddove archeologia e storia si fermano.

In generale, la variabilità genetica viene misurata attraverso l'analisi di porzioni polimorfiche del DNA, che possono variare da un individuo all'altro, mentre la loro frequenza e distribuzione possono essere diverse in differenti popolazioni. Il potere di "risolvere" la diversità fra differenti popolazioni e di evidenziare un'eventuale sub-

struttura all'interno di una determinata popolazione dipende dal sistema genetico analizzato. Oltre alle proprietà del singolo marcatore, sono fondamentali anche le modalità di trasmissione della porzione di DNA su cui esso è localizzato. In particolare, i marcatori polimorfici localizzati rispettivamente sui cromosomi autosomici, sui cromosomi del sesso e sul DNA mitocondriale, hanno un diverso potere nel rivelare il differenziamento genetico fra gli individui e fra le popolazioni. Un uso combinato di questi marcatori rappresenta certamente l'approccio più informativo.

Precedenti studi, condotti su varianti autosomiche, indicano l'assenza di una significativa eterogeneità su larga scala geografica tra le zone costiere e la parte più interna dell'Isola. Storicamente, le prime sono state caratterizzate, nel corso del tempo, da ripetute invasioni ad opera di popoli stranieri, mentre la parte più interna

dell'Isola ne è rimasta sostanzialmente immune [10,11]. Tali evidenze sono state ottenute per il tramite di varianti principalmente dislocate sulla regione HLA. Quest'ultima regione risulta particolarmente informativa nell'analisi genetica, in quanto si tratta del sistema più polimorfico esistente nei mammiferi. Tuttavia, questa stessa regione del genoma, potrebbe essere stata soggetta a pressioni selettive, legate al ruolo svolto dalle proteine per le quali essa stessa codifica: la difesa dai più disparati agenti patogeni. Questo rende difficile determinare quanto tali varianti possano essere rappresentative dell'intero genoma e di antichi eventi demografici. Al fine di ottenere un quadro più accurato della struttura genetica della popolazione sarda, in questo lavoro è stata determinata la struttura dei cromosomi Y di un ampio numero di maschi sardi, utilizzando un pannello di polimorfismi a singolo nucleotide (SNPs) neutrali e di short tandem repeats

(STRs), altresì noti come microsatelliti, dislocati nella porzione non ricombinante, oggi definita porzione specifica del maschio (MSY), di tale cromosoma. Infatti, questo cromosoma per il 95% della sua lunghezza non dà luogo a fenomeni di ricombinazione con il suo omologo, il cromosoma X. L'assenza di ricombinazione contribuisce a trasformare il cromosoma Y in un vero e proprio super-locus genetico, capace di riflettere la storia dei maschi, conservando così una traccia del cammino evolutivo percorso dalla nostra specie dalle origini fino ad oggi. Inoltre, il suo stato aploide, la sua assenza nelle femmine e la sua trasmissione paterlineare hanno importanti conseguenze nella genetica di popolazione. Per queste caratteristiche il cromosoma Y si configura come il sistema più sensibile alla deriva genica la quale, di per sé, rappresenta un'ulteriore fonte di variabilità.

I marcatori testati sono assai utili ai fini delle analisi comparative poiché essi non sono esplicitamente attaccati dalle forze evoluzionistiche legate alla risposta immune contro gli agenti patogeni. La natura non ricombinante di questa porzione del cromosoma Y, così come la disponibilità di un orologio molecolare affidabile, “tarato” su tassi di mutazione definiti degli STRs, conduce ad una ricostruzione cronologica della genealogia degli aplotipi Y, quando questi differenti tipi di marcatori vengano analizzati congiuntamente [12-15].

Alcuni lavori, basati sullo studio di questo cromosoma, hanno riportato un certo grado di micro-eterogeneità nella popolazione sarda, soprattutto in conseguenza del confronto tra aree ad estensione piuttosto limitata e villaggi isolati [16-19].

Da quanto appena esposto si evince come l'interpretazione del panorama genetico sardo nel suo complesso sia tutt'altro che semplice.

Lo scopo di questo studio è quello di stabilire la distribuzione intra-regionale di questi marcatori molecolari sul cromosoma Y e di dedurre il popolamento passato e gli eventi demografici che hanno caratterizzato la Sardegna, al fine di valutare l'impatto di questi fattori negli studi di associazione per malattie multifattoriali.

MATERIALI E METODI

Selezione dei campioni

Come appare chiaro dagli intenti di questo studio, la selezione dei campioni ha rappresentato un momento cruciale.

Sono stati raccolti 187 campioni provenienti dai nidi degli Ospedali Brotzu e San Giovanni di Dio, nonché da quelli delle Cliniche Sant'Elena e Sant'Anna di Cagliari, 103 provenienti dal nido dell'Ospedale San Camillo di Sorgono e 86 dall'Ospedale Civile di Tempio. I campioni di Cagliari comprendono neonati della zona del Campidano, area densamente popolata del sud dell'Isola. Il campione di Sorgono include neonati nell'area della Barbagia, che rappresenta la parte più interna dell'Isola la quale, fino a pochi anni fa, era caratterizzata da un certo isolamento culturale oltre che geografico. Infine, il campione di

Tempio comprende neonati provenienti dalla Gallura, una regione differenziata dal resto dell'Isola, nella quale si parla un dialetto che mostra alti livelli di similitudine con quello parlato nella vicina Corsica.

La selezione di queste tre regioni è stata effettuata al fine di massimizzare la possibilità di trovare eterogeneità e di confrontare regioni con una storia demografica chiaramente differente.

Inoltre, in questo studio è stato incluso un ulteriore campione di maschi sardi, di cui 155 donatori di sangue, provenienti dal Centro Trasfusionale dell'Ospedale Brotzu di Cagliari e 399 pazienti affetti da diabete di tipo 1, già impiegati in un precedente lavoro [20].

Il DNA dei neonati impiegati in questo studio è stato ottenuto mediante l'estrazione con Chelex [21] a partire dalle "Guthrie Cards" utilizzate per gli screening neonatali.

Al fine di includere soli individui di origine sarda, è stato

utilizzato un metodo di selezione basato sui cognomi dei neonati, traendo vantaggio dal fatto che il cognome e il cromosoma Y hanno una trasmissione patrilineare. Utilizzando il sito www.gens.labo.net è stata valutata la distribuzione del cognome di ciascun individuo inserito nello studio in Sardegna e nella Penisola. Solo i cognomi più comuni nell'Isola che non nel resto d'Italia, sono stati inclusi in questo studio, viceversa, sono stati scartati tutti quegli individui il cui cognome era maggiormente diffuso nell'Italia peninsulare che non in Sardegna, indicativi di un flusso genetico recente. Tale selezione è stata semplificata dal fatto che i sardi esibiscono cognomi particolari, chiaramente distinguibili da quelli del resto d'Italia; ciò è dovuto al fatto che la lingua sarda, pur appartenendo al gruppo neolatino delle lingue indoeuropee, risulta assai differenziata dagli altri membri di questa famiglia.

Utilizzando questo sistema di selezione, è stato eliminato solo lo 0,9% dei campioni inizialmente raccolti per questo studio.

Al fine di garantire l'indipendenza dei soggetti inclusi, la selezione dei campioni è stata limitata ad individui nati in un arco temporale di soli otto mesi e, nel caso in cui fossero presenti dei gemelli, è stato inserito un solo individuo.

Genotipizzazione

Single nucleotide polymorphisms (SNPs)

Per la genotipizzazione degli SNPs è stato utilizzato un approccio sequenziale basato sulle frequenze relative dei vari marcatori ottenute da un precedente studio [20].

L'assenza di ricombinazione e il conseguente completo *linkage disequilibrium* tra un dato polimorfismo e un altro, occorso in precedenza sulla stessa linea cromosomica, ha permesso di procedere con uno screening gerarchico di polimorfismi, come mostrato in Figura 1.

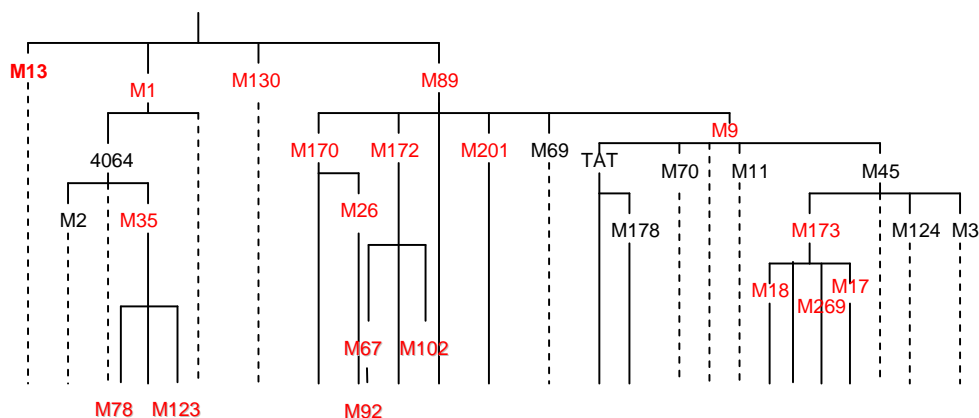


Figura 1. Albero delle mutazioni. In rosso sono indicate le mutazioni genotipizzate in questo studio.

Complessivamente, i campioni raccolti sono stati testati per i seguenti SNPs: M9, M13, M17, M18, M26, M35, M67, M68, M78, M89, M92, M102, M123, M130, M170, M172, M173, M201 M267 e M269. L'inserzione Alu (YAP), è stata impiegata per la genotipizzazione di M1, per un totale di 21 polimorfismi biallelici, dislocati nella porzione specifica del maschio del cromosoma Y[12,13,20,22]. Le genealogie definite dagli SNPs trovati sono indicate in accordo con la

denominazione convenzionale del Y-Chromosome Consortium (2002) [23].

Le mutazioni M1, M9, M26, M35, M89, M130, M170, M172, M173, M201, ed sono state genotipizzate secondo quanto descritto da Contu e colleghi [20].

Le mutazioni, tipizzate ex novo nel presente studio, M17, M18, M67, M78, M123 ed M267 sono state genotipizzate tramite sequenziamento su sequenziatore capillare ABI PRISM 3100 Genetic Analyzer. Le mutazioni M92 ed M102 sono state genotipizzate tramite ARMS PCR, mentre per M13 ed M130 è stato utilizzato il DOT BLOT. In tutti questi casi per le PCR sono state impiegate coppie di primers riportate nel lavoro di Underhill e colleghi [24]. La mutazione M269 è stata infine genotipizzata con il DOT BLOT. I primers e le sonde impiegati per questo scopo sono di seguito riportati:

5'-TCTCCCTTTGCAAAATGTCC

3'-AGGAGATCGTGACCATCCAG

M269 C-gtttggttaat**C**ctggtaaattg TM 58°C

M269 T-gtttggttaat**T**ctggtaaattg TM 56°C

Short Tandem Repeats (STRS)

I 9 STRs considerati in questo studio, *DYS19*, *DYS385a*, *DYS385b*, *DYS389-I*, *DYS389-II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, sono stati genotipizzati impiegando il kit Men Type Argus Y-HM PCR Amplification, in accordo con le raccomandazioni dell'*International Forensic Y-User Group*. Tutti questi STRs sono tetranucleotide-repeats, con l'eccezione del *DYS392* che è un tri-nucleotide repeat. Gli STRs sono stati genotipizzati su sequenziatore capillare ABI PRISM 3100 Genetic Analyzer, come previsto dalle istruzioni del kit.

E' importante sottolineare che 3 dei 9 STRs originariamente tipizzati sono stati esclusi in quanto 2 di questi, *DYS385a* e *DYS385b*, erano duplicati e co-amplificati dalla stessa coppia di primers e, come tali, filogeneticamente indistinguibili, mentre l'altro, il *DYS392*,

ha mostrato un livello di successo di genotipizzazione inferiore alla soglia del 90% fissata per questo studio.

Analisi Statistica

Allo scopo di misurare le distanze genetiche tra coppie di popolazioni sarde è stata impiegata la statistica F_{ST} (Fixation Index-Statistics) per marcatori biallelici [25-27]. In questo modo è stato possibile testare la distribuzione delle frequenze alplotipiche sotto l'ipotesi nulla di “assenza di differenziamento”.

Sugli STRs, che evolvono con un modello *stepwise*, è stato, invece, usato il metodo R_{ST} [26].

La *diversità genetica* e la sua varianza sono state calcolate in accordo con Nei [28]. Tutte e tre le misure, F_{ST} e R_{ST} e *diversità genetica*, sono state determinate con il software Arlequin 3.11 [29]. Le distanze tra F_{ST} e R_{ST} sono state raggruppate usando PHYLIP (PHYlogeny Inference Package) v3.66 tramite il metodo “UPMGA tree drawing method”. Gli alberi senza radice sono stati dunque

disegnati con il programma PHYLIP executive Drowtree [30].

I parametri genetici e demografici sono stati dedotti con un approccio di tipo Bayesiano, mediante il programma BATWING (Bayesian Analysis of Trees With Internal Node Generation), sulla base dell'algoritmo presentato da Wilson and Balding [31]. BATWING utilizza la procedura di Markov Chain Monte Carlo per generare sequenze di alberi genealogici e popolazioni, consistenti con i dati genetici osservati in un campione di individui, sulla base di determinati parametri di partenza che vengono stabiliti e assegnati anticipatamente, i *priors*.

E' stato assegnato un unico *prior* per il tasso di mutazione, basato sulla stima riportata da Zhivotovsky e colleghi [32] pari a $6,9 \times 10^{-4}$ e applicato a DYS19, DYS389I, DYS389II, DYS390, DYS391 and DYS393, mentre *priors* debolmente informativi sono stati assegnati alla taglia

effettiva della popolazione ancestrale e al tasso di crescita della popolazione stessa. Le distribuzioni *posterior*, ottenute con i parametri considerati, si presentavano asimmetriche e non normali. Si è osservato che un'analisi più robusta poteva essere ottenuta a partire dalla moda, anziché dalle più comuni media e mediana [33] e, così, si è scelto di utilizzare la suddetta opzione.

RISULTATI

In questo lavoro è stata stabilita la distribuzione degli aplogruppi del cromosoma Y in 376 neonati maschi sardi (Figura 2 e Materiali e Metodi).

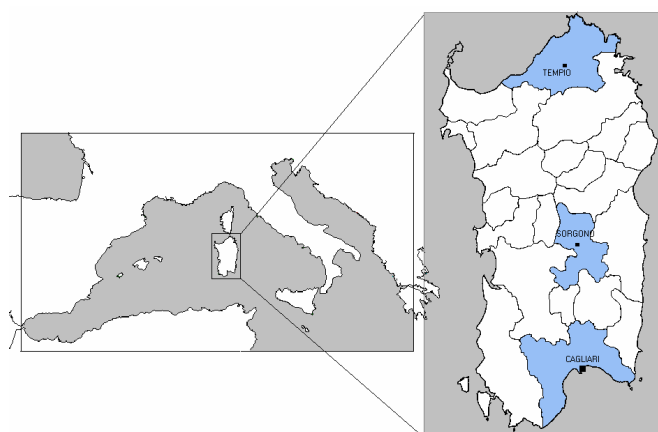


Figura 2. La Sardegna con le tre regioni nelle quali è stato effettuato il campionamento.

La distribuzione dei vari aplogruppi è risultata relativamente omogenea nelle tre differenti aree dell'Isola nelle quali è stato effettuato il campionamento (Tabella 1).

POPOLAZIONE:		CAGLIARI		SORGONO		TEMPIO		
Diversità genetica:		0.828 +/- 0.019		0.775 +/- 0.031		0.821 +/- 0.022		
APLOTIPO		N	Frequenza	N	Frequenza	N	Frequenza	
M13		1	0.005	0	0.000	0	0.000	
M1		1	0.005	0	0.000	0	0.000	
M35	(xM78,M123)	1	0.005	1	0.010	1	0.012	
	M78	12	0.064	3	0.029	3	0.035	
	M123	4	0.021	1	0.010	4	0.047	
	ND	3	0.016	3	0.029	3	0.035	
M89	(xM170,M172,M201,M9,M267)	0	0.000	0	0.000	3	0.035	
	(xM170,M172,M201,M9,ND)	2	0.011	1	0.010	0	0.000	
	M267	5	0.027	3	0.029	0	0.000	
	M170	(xM26)	6	0.032	0	0.000	0	0.000
		M26	59	0.316	38	0.369	24	0.279
	M172	(xM102,M92)	5	0.027	0	0.000	2	0.023
		M102	5	0.027	3	0.029	0	0.000
		M92(xM67)	3	0.016	0	0.000	1	0.012
		M92,M67	7	0.037	3	0.029	3	0.035
		ND	2	0.011	0	0.000	1	0.012
	M201	26	0.139	12	0.117	18	0.209	
	M9	(xM173)	2	0.011	2	0.019	6	0.070
		M173(xM17,M18,M269)	4	0.021	0	0.000	0	0.000
		M173,M17	1	0.005	3	0.029	0	0.000
		M173,M18	3	0.016	5	0.049	0	0.000
M173,M269		33	0.176	22	0.214	17	0.198	
M173(xM17,M18,ND)		2	0.011	3	0.029	0	0.000	
TOTALE		187	1.000	103	1.000	86	1.000	

Tabella 1. *Diversità genetica* e frequenze aplotipiche delle genealogie di cromosomi Y nei 376 neonati sardi. **ND:** non determinato

Infatti, gli F_{ST} a coppie, calcolati dalle frequenze degli aplogruppi non mostrano differenze significative per

quanto attiene alla differenziazione tra sotto-popolazioni (Tabella 2a).

	Cagliari	Sorgono	Tempio
Cagliari	*	0.393 ± 0.005	0.269 ± 0.003
Sorgono	0.000	*	0.109 ± 0.003
Tempio	0.002	0.009	*

Tabella 2a. F_{ST} a coppie. I valori di F_{ST} sono mostrati sotto la diagonale. I corrispondenti valori di P con un livello di significatività $P=0.05$ sono mostrati sopra la diagonale.

In modo analogo, gli R_{ST} a coppie, calcolati sullo stesso campione, utilizzando i marcatori STR, non ha mostrato evidenza significativa di differenziazione tra sotto-popolazioni (Tabella 2b).

	Cagliari	Sorgono	Tempio
Cagliari	*	0.310 ± 0.005	0.898 ± 0.003
Sorgono	0.001	*	0.212 ± 0.004
Tempio	-0.007	0.005	*

Tabella 2b. R_{ST} a coppie. I valori di R_{ST} sono mostrati sotto la diagonale. I corrispondenti valori di P con un livello di significatività $P=0.05$ sono mostrati sopra la diagonale.

Al fine di massimizzare la possibilità di valutare l’impatto della presenza di un’eventuale sub-struttura negli studi di associazione, le frequenze delle varianti bialleliche di un campione di 399 pazienti affetti da diabete di tipo 1, provenienti dal sud dell’Isola e genotipizzati in un precedente studio [20], sono state confrontate con quelle del *pool* di 376 neonati raccolti nelle zone del Campidano, della Barbagia e della Gallura. Anche sotto questo scenario forzato, assai differente da quello che si verificherebbe nel caso di uno studio caso-controllo, in cui

casi e controlli vengono raccolti in simili proporzioni da differenti sub-regioni, non si è ottenuta alcuna evidenza di eterogeneità ($F_{ST}=0.001$, $P=0.157$).

Poiché da quanto appena esposto, risulta evidente che la popolazione sarda si caratterizza per un elevato grado di omogeneità, i 376 neonati, i 399 pazienti diabetici ed un campione addizionale di 155 donatori di sangue, per un totale di 930 maschi sardi sono stati impiegati per stabilire le frequenze delle genealogie di cromosomi Y sulla base di 19 marcatori biallelici (Tabella 3).

Diversità genetica:		0,8013 +/- 0,0101		
APLOTIPO		N	Frequenza	
M13		1	0.00	
M1		5	0.01	
M35	(xM78,M123)	9	0.01	
	M78	41	0.04	
	M123	19	0.02	
	ND	9	0.01	
M89	(xM170,M172,M201,M9,M267)	12	0.01	
	(xM170,M172,M201,M9,ND)	4	0.00	
	M267	21	0.02	
	M170	(xM26)	23	0.02
		M26	344	0.37
	M172	(xM102,M92)	27	0.03
		M102	22	0.02
		M92(xM67)	16	0.02
		M92,M67	26	0.03
		ND	3	0.00
	M201	117	0.13	
	M9	(xM173)	22	0.02
		M173(xM17,M18,M269)	13	0.01
M173,M17		13	0.01	
M173,M18		16	0.02	
M173,M269		158	0.17	
M173(xM17,M18,ND)		9	0.01	
TOTALE		930	1.00	

Tabella 3. *Diversità genetica* e frequenze aplotipiche delle genealogie di cromosomi Y nei 376 neonati sardi. **ND:** non determinato.

Le mutazioni M130 ed M68 sono state testate, ma non sono state trovate nel campione considerato.

Le frequenze degli aplogruppi osservati nel campione di 930 maschi sardi sono state, dunque, confrontate con quelle riportate per altre popolazioni europee. Tale confronto separa nettamente la popolazione sarda dal restante gruppo europeo [12,14,34,35] (Figura 3).

L'aplogruppo principalmente responsabile del differenziamento dei sardi rispetto alle altre popolazioni europee è certamente l'I-M26. Questa variante mostra, infatti, una frequenza complessiva pari allo 0.37 in Sardegna, mentre è virtualmente assente nelle altre popolazioni, inclusa la vicina Corsica [34]. Questo aplogruppo è stato ritrovato, a frequenze assai minori, solo nei Baschi [14,36] e in poche altre popolazioni occidentali europee [37-39].

In aggiunta, il raro SNP M18 sembra essere rappresentato unicamente nella popolazione sarda. Inoltre, alcuni aplogruppi presenti e relativamente frequenti nell'Isola, quali ad esempio G-M201 ed E-M78, sono stati ritrovati solo nei paesi a Est dell'Europa, mentre un altro aplogruppo, R-M269, comune in Sardegna, è assai più comune nelle popolazioni dell'Ovest europeo [12].

Il risultato di questa distribuzione di genealogie di cromosomi Y in Sardegna è un elevato grado di *diversità genetica* (0.801 ± 0.010) (Tabella 3) ed una significativa differenziazione dagli altri gruppi etnici per i quali i dati genetici sono attualmente disponibili (Tabella 4).

Tabella 4a

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.000													
2	-0.000	0.000												
3	-0.001	0.006	0.000											
4	0.078	0.106	0.070	0.000										
5	0.147	0.161	0.157	0.129	0.000									
6	0.076	0.107	0.078	0.010	0.091	0.000								
7	0.087	0.122	0.095	0.029	0.145	0.031	0.000							
8	0.079	0.106	0.084	0.033	0.095	0.024	-0.001	0.000						
9	0.161	0.169	0.176	0.159	0.015	0.132	0.175	0.113	0.000					
10	0.222	0.233	0.240	0.214	0.024	0.201	0.257	0.190	0.009	0.000				
11	0.237	0.256	0.251	0.184	0.300	0.233	0.164	0.167	0.334	0.406	0.000			
12	0.190	0.217	0.203	0.138	0.260	0.178	0.082	0.127	0.306	0.369	0.072	0.000		
13	0.292	0.312	0.336	0.287	0.106	0.332	0.388	0.296	0.068	0.018	0.523	0.494	0.000	
14	0.222	0.240	0.229	0.161	0.325	0.210	0.146	0.160	0.363	0.439	0.008	0.080	0.559	0.000

Tabella 4b

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*													
2	0.414+-0.005	*												
3	0.442+-0.005	0.146+-0.003	*											
4	0.000+-0.000	0.000+-0.000	0.000+-0.000	*										
5	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*									
6	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.069+-0.002	0.001+-0.000	*								
7	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.002+-0.001	0.000+-0.000	0.020+-0.001	*							
8	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.032+-0.002	0.416+-0.005	*						
9	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.102+-0.003	0.000+-0.000	0.000+-0.000	0.000+-0.000	*					
10	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.090+-0.003	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.187+-0.004	*				
11	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*			
12	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.004+-0.000	*		
13	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.002	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.001+-0.000	0.109+-0.003	0.000+-0.000	0.000+-0.000	*	
14	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.219+-0.004	0.002+-0.000	0.000+-0.000	*

Tabella 4. 4a: FST a coppie tra diverse popolazioni europee. **4b:** P values dell’FST.

1.Cagliari, **2.**Sorgono, **3.**Tempio, **4.**Anatolia, **5.**Italia centrale, **6.**Sicilia, **7.**Albania, **8.**Grecia, **9.**Andalusia, **10.**Catalogna, **11.**Polonia, **12.**Croazia, **13.**Paesi Baschi, **14.**Ucraina. Le frequenze geniche impiegate per l’analisi i cui risultati sono mostrati in tabella sono state desunte da Semino e colleghi, Bosch e colleghi, Francalacci e colleghi e Cinnioglu e colleghi[12,14,34,35].

Un albero costruito usando le distanze F_{ST} e il metodo UPGMA, illustra graficamente l'elevato livello di differenziazione dei sardi rispetto alle altre popolazioni (Figura 3a).

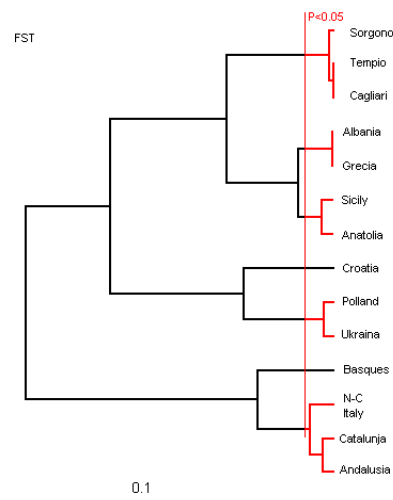


Figura 3. Confronto della popolazione sarda con le altre popolazioni europee basato sull' F_{ST} .

Una possibile riserva nei confronti di questo scenario, è data dal fatto che il campione originariamente impiegato per lo screening degli SNPs del cromosoma Y [13]

includeva anche campioni sardi. In linea di principio, questo tipo di *bias* di accertamento potrebbe incrementare il differenziamento dei sardi rispetto alle altre popolazioni. Comunque, risultati consistenti sono stati ottenuti anche calcolando gli R_{ST} a coppie (Tabella 5 e Figura 4) basati esclusivamente sui dati derivanti dagli STR, i quali non sono soggetti a questo tipo di *bias*.

Tabella 5a

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.000													
2	0.001	0.000												
3	-0.007	0.005	0.000											
4	0.081	0.111	0.071	0.000										
5	0.114	0.132	0.100	0.058	0.000									
6	0.071	0.110	0.063	0.024	0.068	0.000								
7	0.212	0.245	0.197	0.073	0.042	0.128	0.000							
8	0.119	0.134	0.110	0.067	0.012	0.092	0.051	0.000						
9	0.075	0.100	0.058	0.005	0.033	0.040	0.048	0.035	0.000					
10	0.107	0.124	0.095	0.063	0.008	0.082	0.058	0.004	0.024	0.000				
11	0.197	0.202	0.182	0.176	0.239	0.271	0.224	0.226	0.150	0.227	0.000			
12	0.159	0.184	0.138	0.095	0.170	0.184	0.151	0.172	0.069	0.158	0.034	0.000		
13	0.118	0.119	0.112	0.102	0.022	0.131	0.098	0.018	0.060	0.011	0.227	0.185	0.000	
14	0.238	0.251	0.223	0.200	0.280	0.306	0.279	0.292	0.175	0.278	-0.001	0.036	0.291	0.000

Tabella 5b

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*													
2	0.310+-0.004	*												
3	0.898+-0.003	0.217+-0.004	*											
4	0.000+-0.000	0.000+-0.000	0.000+-0.000	*										
5	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*									
6	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*								
7	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*							
8	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.005+-0.000	0.000+-0.000	0.000+-0.000	*						
9	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.165+-0.004	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*					
10	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.018+-0.001	0.000+-0.000	0.000+-0.000	0.104+-0.003	0.000+-0.000	*				
11	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*			
12	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	*		
13	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.001+-0.000	0.000+-0.000	0.000+-0.000	0.002+-0.000	0.000+-0.000	0.013+-0.001	0.000+-0.000	0.000+-0.000	*	
14	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.000+-0.000	0.453+-0.005	0.001+-0.000	0.000+-0.000	*

Tabella 5. 5a: RST a coppie tra diverse popolazioni europee. **5b:** P values dell'RST.

1.Cagliari, **2.**Sorgono, **3.**Tempio, **4.**Anatolia, **5.**Italia centrale, **6.**Sicilia, **7.**Albania, **8.**Grecia, **9.**Andalusia, **10.**Catalogna, **11.**Polonia, **12.**Croazia, **13.**Paesi Baschi, **14.**Ucraina. Le frequenze geniche impiegate per l'analisi i cui risultati sono mostrati in tabella sono state desunte da Roewer e colleghi [40]. I dati relativi al campione sardo sono originali.

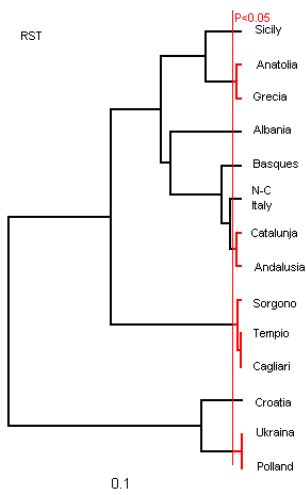


Figura 4. Confronto della popolazione sarda con le altre popolazioni europee basato sull' R_{ST} .

Successivamente, è stata effettuata l'analisi con BATWING, al fine di stabilire il TMRCA individuale (Time of Most Recent Common Ancestor) per gli aplogruppi più rappresentativi (I-M26, R-M269, G-M201, E-M78) del campione sardo. Essi rendono conto, complessivamente, del 71% degli aplogruppi Y presenti in Sardegna. Tali aplogruppi, sono stati analizzati unitamente ai 6 marcatori STR. I valori di TMRCA per questi aplogruppi sono compresi tra 19.500 e 22.800 anni fa (Tabella 6).

APLOGRUPPO Y	TMRCA	95% c.i.
I-M26	21.0	16.0-25.5
R-M269	22.8	17.0-25.4
G-M201	19.5	14.0-22.2
E-M78	21.0	19.4-30.8
R-M18	8.1	7.8-10.5

Tabella 6. Valori di TMRCA dei principali aplogruppi presenti nella popolazione sarda ottenuti tramite l'analisi con BATWING.

Inoltre, l'analisi è stata estesa ad un aplogruppo raro e "sardo-specifico", R-M18, per il quale il valore di TMRCA è risultato essere pari a 8.100 anni fa (7.800-10.500 anni fa).

L'analisi con BATWING è stata, inoltre, impiegata per analizzare i dati provenienti dai 376 neonati sardi al fine di valutare quando la popolazione di quest'isola ha iniziato ad espandersi. Il valore così ottenuto è pari a 14.000 anni

fa (7.800-22.000 anni). Al fine di inserire questi dati nel contesto di quelli osservati nelle altre popolazioni europee, sono stati impiegati dati già pubblicati [14,35,41] ed è stato calcolato il tempo di espansione iniziale delle popolazioni di interesse. Per quanto attiene all'Anatolia, il valore trovato è pari a 9.500 anni fa (6.700-13.700), mentre per le popolazioni Balcanica, da una parte, e Iberica, dall'altra, sono stati ottenuti i seguenti risultati: 7.400 anni fa (6.500-16.100) e 9.800 anni fa (9.500-13.700) rispettivamente. L'analisi con BATWING suggerisce inoltre che l'effettiva taglia della popolazione all'inizio dell'espansione fosse pari a 789 (450-1264) individui per quanto concerne i sardi, 1741 (957-2857) per gli anatolici e 601 (218-867) per gli abitanti dei Balcani e 339 (273-668) per gli iberici.

DISCUSSIONE

I nostri risultati indicano che le tre aree della Sardegna studiate nel presente lavoro non esibiscono alcuna significativa evidenza di eterogeneità genetica interregionale, qualunque sia il parametro di misura della differenziazione impiegato. In particolare, le subpopolazioni provenienti dalle regioni costiere (Campidano e Gallura), le quali nel corso del tempo sono state sottoposte a svariate dominazioni culturali e politiche, non differiscono in modo statisticamente significativo dalla parte più interna e isolata della Sardegna (Barbagia), la quale storicamente non ha mai subito invasioni da parte di popoli stranieri. Ciò, in pieno accordo con altri studi che hanno coinvolto differenti cromosomi e campionamenti indipendenti [10,11].

La distribuzione omogenea di varianti in diverse macro-aree dell'Isola è consistente con una comune antica origine e con un comune substrato culturale. Per esempio, i Nuraghe, le tipiche torri preistoriche, sono ampiamente e uniformemente diffusi attraverso tutta l'Isola.

I dati derivanti da questo studio, inoltre, offrono qualche ulteriore delucidazione riguardo al periodo di fondazione della popolazione sarda. Infatti, gli aplogruppi del cromosoma Y più comunemente rappresentati in Sardegna mostrano un TMRCA compreso tra 19.500 e 22.800 anni, indipendentemente dalla loro genealogia, il che è indicativo di un effetto fondatore piuttosto antico. In particolare, l'aplogruppo I-M26, che mostra un TMRCA di 21.000 anni (16.000-25.500 anni), è particolarmente informativo, in quanto è presente in circa il 37% dei sardi, mentre è molto raro negli altri gruppi umani. Inoltre I-M26 rappresenta un ramo terminale dell'albero che contiene le mutazioni a

carico della regione MSY, il che evidenzia che si tratta di un aplogruppo relativamente recente, per cui è improbabile che le linee I-M26 presenti nelle tribù dei primi fondatori sardi presentassero un'elevata variabilità.

In altri termini, è verosimile che il grosso della variabilità intra-aplogruppo, per I-M26, presente oggi in Sardegna si sia generata nei millenni dopo la fondazione della popolazione sarda e che il suo valore di TMRCA sia dunque correlato con l'iniziale occupazione dell'Isola e non con eventi demografici precedenti rispetto a tale fondazione.

Tra l'altro l'analisi con BATWING rivela che la popolazione maschile sarda ha cominciato ad espandersi all'incirca 14.000 anni fa, con una taglia iniziale della popolazione dell'ordine di un migliaio di individui. Anche in questo caso, il segnale dell'espansione rilevato da BATWING probabilmente non predata l'arrivo dei sardi fondatori

nell'isola perché si è osservato che gli altri campionamenti europei forniscono un più recente tempo di espansione iniziale.

Così, nonostante il modello di demografia semplificato implicito in questo tipo di calcoli e gli intervalli di confidenza stimati da BATWING, un quadro coerente mostra un'espansione demografica della popolazione sarda che risale ad un'epoca pre-neolitica. Tra l'altro, i valori desunti da BATWING sono conservativi, utilizzando un tempo di generazione di 25 anni.

Inoltre, per ottenere un risultato più robusto e affidabile, è stato utilizzato un approccio basato sulla moda in luogo di quello più comunemente impiegato, che si basa sulle misure della media, che tende a fornire valori assai più antichi per qualunque parametro valutato con BATWING (Materiali e Metodi e risultati non pubblicati).

Tali stime genetiche, in accordo con i dati archeologici [42,43], suggeriscono che la Sardegna fosse già popolata in un periodo di tempo compreso tra 18.000 e 10.000 anni fa. Inoltre, i dati genetici sono consistenti con quelli geologici che evidenziano una connessione tra la Sardegna e l'Italia peninsulare (tramite la Corsica) durante l'ultimo periodo glaciale. Tale "ponte" potrebbe aver consentito il passaggio di persone per diverse migliaia di anni, fino a circa 10.000 anni fa. Le evidenze di un precoce popolamento neolitico, provengono anche da altri marcatori genetici, quali gli aplogruppi mitocondriali V e H3 [44,45]. La Sardegna potrebbe aver attratto cacciatori-raccoglitori paleolitici, grazie al suo clima temperato, alla presenza animali di grossa taglia quali, *Megaceros cazioti* e grazie alla presenza di altre specie endemiche caratterizzate da un elevato tasso di riproduttività, ed altri tipi di cibo, quali ad esempio frutti provenienti dalle

foreste di cui l'Isola era ricoperta o mitili presenti nei laghi salati presenti nelle regioni rivierasche.

L'assenza di I-M26 nella vicina Corsica [34], risulta quantomeno sorprendente. Infatti, in base al nostro modello favorito di popolamento della Sardegna, la Corsica è stata attraversata dagli antichi sardi durante gli eventi di popolamento iniziale e I-M26 avrebbe dovuto lasciare traccia di questo passaggio sotto forma di gradiente di frequenza. Comunque, fonti storiche suggeriscono che la Corsica sia stata ampiamente ripopolata nel Medioevo, così da cancellare le tracce di un'eventuale presenza di I-M26, come se il ri-popolamento avesse mascherato il substrato genetico ancestrale di quest'isola.

I dati e le conclusioni presentate in questo studio, in modo particolare l'assenza di una significativa sub-struttura nella popolazione sarda, hanno un impatto considerevole negli studi di associazione estesi all'intero genoma. Per

questa popolazione, infatti, si configura la possibilità di effettuare studi di associazione con disegno caso-controllo estesi a tutto il genoma, caratterizzati da solidità sperimentale e grande potere statistico.

Altre popolazioni fondatrici, quali ad esempio quelle finlandesi e islandesi, mostrano una qualche sub-struttura confinata ad aree geograficamente distinte [8,46]. Esistono anche in Sardegna, indicazioni di micro-eterogeneità in sub-isolati, che però appaiono chiaramente effetti secondari. Per esempio, uno studio avrebbe suggerito la presenza di eterogeneità attraverso l'analisi di 21 sub-regioni dell'Isola, suddivise sulla base di criteri storico-linguistici. [18]. Tuttavia, un'analisi della varianza molecolare (AMOVA) effettuata impiegando le frequenze riportate dagli autori, indicano che la maggior parte della variabilità si trova all'interno delle sub-popolazioni testate

(99,8%), mentre meno dello 0,2% di tale variabilità si trova tra popolazioni (Tabella 7).

MARCATORE	F _{ST}	P(rand.value ≥ obs. value)	PERCENTUALE DI VARIABILITA'	
			TRA POPOLAZIONI	ALL'INTERNO DELLE POPOLAZIONI
ABO	0.000	0.575+-0.006	-0.04	100.04
ACP1	0.002	0.102+-0.003	0.15	99.85
AK1	0.003	0.062+-0.002	0.15	99.85
C3	0.002	0.140+-0.003	0.24	99.76
CDE	0.006	0.000+-0.000	0.56	99.44
DIA	0.002	0.096+-0.003	0.23	99.77
ESD	0.004	0.014+-0.001	0.36	99.64
GC	0.001	0.311+-0.005	0.08	99.92
KEL	-0.002	0.780+-0.005	-0.19	100.19
MN	0.002	0.206+-0.004	0.18	99.82
PGM1	0.001	0.217+-0.004	0.11	99.89
6PGD	0.001	0.314+-0.005	0.06	99.94
MEDIA	0.002		0.16	99.84

Tabella 7. AMOVA su 12 marcatori pre-molecolari su 21 domini linguistici della Sardegna, calcolati sulle frequenze alleliche riportate in Tabella 1a pagina 130 di Cappello e colleghi [18].

Tra l'altro, gli stessi criteri suddivisione delle sub-regioni mostrano scarso fondamento, in quanto la lingua sarda e i

suoi dialetti appaiono assai più recenti degli eventi demografici di popolamento che hanno interessato l'Isola.

Un altro studio avrebbe trovato un qualche grado di eterogeneità nella distribuzione di marcatori del cromosoma Y di individui con cognomi monofiletici provenienti da differenti sub-regioni dell'Isola [16], ma, consistentemente con i risultati ottenuti nel presente studio, non è stata trovata alcuna evidenza di eterogeneità significativa, quando gli individui genotipizzati sono stati analizzati in funzione del solo luogo di nascita. Infatti, la sola selezione del cognome monofiletico, ha determinato la perdita di una rilevante e condivisa parte di variabilità, che è, così, stata rimossa dalla successive analisi. Tale selezione dei cognomi monofiletici è anche più prona all'effetto cumulativo legato alla mispaternità.

Una debole evidenza di eterogeneità nella distribuzione dei marcatori del cromosoma Y di individui provenienti da

varie sub-regioni sarde, è stata riportata anche da Scozzari e colleghi [17], ma solo per uno dei 17 marcatori testati e considerando un numero estremamente limitato di campioni (appena 18 individui in una delle aree considerate). Inoltre, non va sottovalutato il fatto che lo studio in questione presentava ulteriori aspetti problematici nel disegno sperimentale in quanto comparava la variabilità genetica in piccoli paesi di poco più di mille abitanti, con quella di province che vantano una popolazione di centinaia di migliaia di individui.

I risultati di tali studi vanno interpretati tenendo conto che molti colli di bottiglia demografici si sono verificati durante il popolamento dei piccoli e dispersi villaggi dell'Isola [19,47,48], i quali possono rendere conto della microeterogeneità e dell'elevato grado del substrato di linkage disequilibrium osservato in queste aree piccole e

isolate, in confronto con la popolazione sarda in generale [48].

Così, questo studio conferma la mancanza di una significativa differenziazione tra macro aree sarde. Ciò è consistente con il fatto che patologie quali il diabete di tipo 1 e la sclerosi multipla non esibiscono alcuna significativa differenza di incidenza tra le differenti sub-regioni sarde.

In aggiunta, mentre nella popolazione sarda è stata osservata una distribuzione dei marcatori assai omogenea, non è stata rilevata alcuna restrizione della variabilità interindividuale. Per esempio, i sardi mostrano un più alto grado di variabilità che non la popolazione spagnola [14,36]. In generale, le varianti del cromosoma Y comuni nelle popolazioni dell'ovest dell'Europa che predatano la fondazione della Sardegna sono riscontrate a frequenze apprezzabili nella popolazione sarda. Questo assicura la possibilità di trovare varianti di malattia comuni sia alle

popolazioni dell'Ovest che a quelle dell'Est dell'Europa. Inoltre, queste caratteristiche, unitamente alla considerazione che la popolazione è piuttosto antica, rende conto delle numerose varianti aploipiche altamente informative, attraverso tutto il genoma. E ciò potrebbe essere particolarmente utile durante la fase di fine mappaggio post-svelamento finalizzata alla dissezione e alla localizzazione di varianti eziologiche individuali nella malattie comuni nell'Isola [49].

I dati derivanti da queste analisi sul cromosoma Y sottolineano chiaramente l'unicità della popolazione sarda e indicano che forti effetti fondatori non sono limitati a tratti monogenici quali ad esempio la Beta talassemia [2] o il morbo di Wilson [3], ma includono anche SNPs neutrali e comuni, come illustrato per la variante I-M26. Questo suggerisce che grazie a tali effetti fondatori, alcune varianti di malattia possano essere uniche nei sardi o che possano

trovarsi in un *range* ottimale di frequenza tale da poter portare ad una loro identificazione in questa popolazione. Complessivamente questi dati forniscono uno scenario assai promettente per gli studi di associazione su tutto il genoma e per i susseguenti sforzi di mappaggio fine finalizzati a trovare polimorfismi eziologici nella ricerca sulle malattie multifattoriali.

BIBLIOGRAFIA

1. Siniscalco M, Bernini L, Latte B, Motulsky AG (1961) Favism and Thalassæmia in Sardinia and their Relationship to Malaria. *Nature* 190: 1179-1180.
2. Rosatelli MC, Dozy A, Faa V, Meloni A, Sardu R, et al. (1992) Molecular characterization of beta-thalassemia in the Sardinian population. *Am J Hum Genet* 50: 422-426.
3. Loudianos G, Dessi V, Lovicu M, Angius A, Figus A, et al. (1999) Molecular characterization of wilson disease in the Sardinian population--evidence of a founder effect. *Hum Mutat* 14: 294-303.
4. Rosatelli MC, Meloni A, Devoto M, Cao A, Scott HS, et al. (1998) A common mutation in Sardinian autoimmune polyendocrinopathy- candidiasis-ectodermal dystrophy patients. *Hum Genet* 103: 428-434.
5. Karvonen M, Tuomilehto J, Libman I, La Porte R (1993) A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. World Health Organization DIAMOND Project Group. *Diabetologia* 36: 883-892.
6. Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, et al. (2006) The epidemiology of multiple sclerosis in Europe. *Eur J Neurol* 13: 700-722.
7. Marrosu MG, Cocco E, Lai M, Spinicci G, Pischedda MP, et al. (2002) Patients with multiple sclerosis and risk of type 1 diabetes mellitus in Sardinia, Italy: a cohort study. *The Lancet* 359: 1461-1465.
8. 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.
9. Lilliu G (1982) *La civiltà nuragica*. Sassari: Delfino.
10. Lampis R, Morelli L, Congia M, Macis MD, Mulargia A, et al. (2000) The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases. *Hum Mol Genet* 9: 2959-2965.

11. Contu L, Arras M, Carcassi C, La Nasa G, Mulargia M (1992) HLA structure of the Sardinian population: a haplotype study of 551 families. *Tissue Antigens* 40: 165-174.
12. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290: 1155-1159.
13. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26: 358-361.
14. Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, et al. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 68: 1019-1029.
15. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* 98: 10244-10249.
16. 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.
17. Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, et al. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62: 871-884.
18. 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 (Pt 2): 125-141.
19. 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.
20. Contu D, Morelli L, Zavattari P, Lampis R, Angius E, et al. (2002) Sex-related bias and exclusion mapping of the nonrecombinant portion of chromosome Y in human type 1 diabetes in the isolated founder population of Sardinia. *Diabetes* 51: 3573-3576.

21. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506-513.
22. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, et al. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7: 996-1005.
23. Y-Chromosome-Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12: 339-348.
24. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, et al. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65: 43-62.
25. Dempster A, Laird N, Rubin D (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39: 1-38.
26. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 12: 921-927.
27. Lange K (1997) *Mathematical and statistical methods for genetic analysis*. New York, NY, USA: Springer.
28. Nei M (1987) *Molecular evolutionary genetics*. New York, NY, USA: Columbia University Press. 512 p.
29. Excoffier L, Laval G, Schneider S (2005) An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.
30. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.
31. Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150: 499-510.
32. Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, et al. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74: 50-61.
33. Hedges SB, Shah P (2003) Comparison of mode estimation methods and application in molecular clock analysis. *BMC Bioinformatics* 4: 31.
34. Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, et al. (2003) Peopling of three Mediterranean islands (Corsica,

- Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthropol* 121: 270-279.
35. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, et al. (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114: 127-148.
 36. Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, et al. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* 12: 855-863.
 37. Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, et al. (2003) A Y chromosome census of the British Isles. *Curr Biol* 13: 979-984.
 38. Bosch E, Calafell F, Rosser ZH, Norby S, Lynnerup N, et al. (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum Genet* 112: 353-363.
 39. Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, et al. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75: 128-137.
 40. Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, et al. (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int* 118: 106-113.
 41. Bosch E, Calafell F, Gonzalez-Neira A, Flaiz C, Mateu E, et al. (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70: 459-487.
 42. Sondaar PY, Elburg R, Klein-Hofmeijer G, Martini F, Sanges M, et al. (1995) The human colonization of Sardinia: a Late-Pleistocene human fossil from Corbeddu cave. *Comptes Rendus de l'Académie des Sciences Paris* 320: 145-150.
 43. Sondaar PY (1998) Palaeolithic Sardinians: paleontological evidence and methods. In: Balmuth MS, Tykot RH, editors. *Sardinian and Aegean chronology*. Oxford: Oxbow Books. pp. 45-51.
 44. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, et al. (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69: 1348-1356.

45. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75: 910-918.
46. Lappalainen T, Koivumaki S, Salmela E, Huoponen K, Sistonen P, et al. (2006) Regional differences among the Finns: a Y-chromosomal perspective. *Gene* 376: 207-215.
47. Angius A, Bebbere D, Petretto E, Falchi M, Forabosco P, et al. (2002) Not all isolates are equal: linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations. *Hum Genet* 111: 9-15.
48. Zavattari P, Lampis R, Mulargia A, Loddo M, Angius E, et al. (2000) Confirmation of the DRB1-DQB1 loci as the major component of IDDM1 in the isolated founder population of Sardinia. *Hum Mol Genet* 9: 2967-2972.
49. Cucca F, Lampis R, Congia M, Angius E, Nutland S, et al. (2001) A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* 10: 2025-2037.