A.D. MDLXII

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE BIOMEDICHE

INDIRIZZO IN FISIOPATOLOGIA MEDICA XXVII CICLO

Susceptibility to complex diseases in Sardinian population explained by Runs of Homozygosity and genomic regions under positive selection

Direttore della scuola:

Prof. Andrea Piana

Relatore:

Prof. Nicola Glorioso

Dottorando:

Dr. Giovanni Fresu

Responsabile di indirizzo:

Prof. Roberto Manetti

A.A. 2013-2014

A MIA MOGLIE

INDICE

Sommario delle attività svolte durante il triennio della scuola di dottorato in Scienze Biomediche

Durante la durata di tre anni del corso di dottorato ho partecipato allo sviluppo di diversi protocolli inerenti la Farmacogenomica Cardiovascolare e la Genetica di Popolazioni: i lavori scientifici relativi a 3 dei protocolli sopra citati sono stati pubblicati su Pharmacogenomics e PLoS One (vedi).

- La linea principale della nostra ricerca è sulla risposta della pressione arteriosa alla terapia con idroclorotiazide (HCTZ). I diuretici tiazidici sono indicati come i farmaci di prima linea per il trattamento antiipertensivo, anche se rimane del tutto empirica la scelta del miglior farmaco ad un dato paziente. Nella nostra ricerca abbiamo individuato, analizzando 343 pazienti sardi e 142 provenienti da Milano, due geni come plausibili candidati alla risposta pressoria in pazienti affetti da ipertensione arteriosa essenziale.
Abbiamo identificato i geni TET2 e CSMD1 come candidati plausibili nell' influenzare la risposta della pressione arteriosa all' idroclorotiazide. La specificità per l'idroclorotiazide è stata verificata su una coorte indipendente di ipertesi essenziali trattati con Losartan.
- Un altro protocollo di ricerca è stato svolto in collaborazione con il gruppo di ricerca del Prof. Nelson Ruiz-Opazo (Boston University School of Medicine) è stato individuato all'interno della popolazione maschile sarda il gene il ATP1A1 come gene legato all'ipertensione. ATP1A1 codifica per 1Na, K-ATPase la pompa unica del sodio situata nell'endotelio vascolare e nelle celule epiteliali del tubulo renale. Dal sequenziamento di questo gene è emersa una sequenza di inserzione/delezione lunga 12 Timidine (12T), nella regione regolatrice. L'inserzione 12T nei maschi sardi portava ad una diminuzione della pressione sistolica(sBP) di 12,1mmHg e della diastolica (dBP) di 6,6mmHg.
- Un ulteriore protocollo di ricerca sulla farmacogenomica dei Beta bloccanti è in fase di elaborazione statistica. E' stata analizzata una corte di n=403 pazienti affetti da ipertensione arteriosa essenziale, mai trattati, da cui è stato prelevato sangue per le valutazione fenotipiche e per l'estrazione del DNA: è stato quindi valutato l'effetto a 4 ed 8 settimane della terapia con beta-bloccanti. Il lavoro nasce dalla poca conoscenza sull' "azione secondaria" antiipertensiva dei beta-bloccanti, che hanno la loro azione primaria sulla frequenza cardiaca attraverso l'inibizione dei recettori B1 adrenergici.
- Da ultimo, ho partecipato allo sviluppo del protocollo scientifico che ha come obiettivo la definizione della "genetica della massa cardiaca": solo circa il 28% dei pazienti affetti da Ipertensione arteriosa essenziale presentano massa cardiaca elevata per cui questo fenotipo, sfavorevole nei confronti degli outcome CV dei pazienti, è ovviamente almeno in parte determinato da "altro" rispetto al puro "effetto idraulico" della pressione nelle arterie. Il protocollo è stato completato ed i dati sono attualmente in fase di elaborazione statistica.
- Il protocollo scientifico oggetto della mia tesi di dottorato é stato condotto in

collaborazione con l'Università di Torino ed è basato sulle caratteristiche genetiche della popolazione sarda al fine di documentarne le caratteristiche estremamente favorevoli alla conduzione di studi genetici relativi alle malattie complesse quali l'ipertensione arteriosa, la cardiopatía ischemica, il diabete mellito, etc.

Lavori pubblicati

1.
Genome-wide association study identifies CAMKID variants involved in blood pressure response to losartan: the SOPHIA study.
Frau F, Zaninello R, Salvi E, Ortu MF, Braga D, Velayutham D, Argiolas G**,** **Fresu G**, Troffa C, Bulla E, Bulla P, Pitzoi S, Piras DA, Glorioso V, Chittani M, Bernini G, Bardini M, Fallo F, Malatino L, Stancanelli B, Regolisti G, Ferri C, Desideri G, Scioli GA, Galletti F, Sciacqua A, Perticone F, Degli Esposti E, Sturani A, Semplicini A, Veglio F, Mulatero P, Williams TA, Lanzani C, Hiltunen TP, Kontula K, Boerwinkle E, Turner ST, Manunta P, Barlassina C, Cusi D, Glorioso N.
Pharmacogenomics. 2014 Sep;15(13):1643-52. doi: 10.2217/pgs.14.119.

2.
Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection.
Di Gaetano C, Fiorito G, Ortu MF, Rosa F, Guarrera S, Pardini B, Cusi D, Frau F, Barlassina C, Troffa C, Argiolas G, Zaninello R, **Fresu G**, Glorioso N, Piazza A, Matullo G.
PLoS One. 2014 Mar 20;9(3):e91237. doi: 10.1371/journal.pone.0091237. eCollection 2014.

3.
Sex-specific effects of NLRP6/AVR and ADM loci on susceptibility to essential hypertension in a Sardinian population.
Glorioso N, Herrera VL, Didishvili T, Ortu MF, Zaninello R, **Fresu G**, Argiolas G, Troffa C, Ruiz-Opazo N.
PLoS One. 2013 Oct 11;8(10):e77562. doi: 10.1371/journal.pone.0077562. eCollection 2013.

Sottomesso per pubblicazione
TET2 and CSMD1 genes affect systolic blood pressure response to hydrochlorothiazide innever treated essential hypertensives",

Lavori in sviluppo
Farmacogenomica dei Beta bloccanti (tiolo da definire)
Fresu G et al
Definizione della "genetica della massa cardiaca (tiolo da definire)
Fresu G et al

Introduzione

La posizione geografica della Sardegna e la montuosità del suo territorio hanno fatto sì che la popolazione sarda (analogamente a quanto è accaduto ad altre popolazioni europee) per via dell'isolamento, dell'endogamia e dell'azione di particolari processi evolutivi quali la deriva genetica sia caratterizzata da peculiarità antropologiche e geniche. Nell'albero filogenetico dell'Europa, la Sardegna viene separata alla seconda fissione, presenta 1.6 milioni di abitanti ed una superficie di 24000 Km^2, situata a circa 200 km sia dalla coste italiane sia da quelle dell'Africa settentrionale. L'Isola è geograficamente molto vicina alla Corsica, ma le due popolazioni, diverse dal punto di vista etnico, hanno avuto una storia profondamente diversa.

I primi abitanti della Sardegna furono uomini preneolitici nella zona centrosettentrionale. Più antichi insediamenti umani risalgono a circa 10000 anni fa[1]. I reperti archeologici più caratteristici e importanti sono i “nuraghe”, abitazioni costruite in pietra (ritenute da alcuni anche fortezze) di forma particolare, non molto diverse da quelle rinvenute in alcune altre isole o zone costiere del bacino mediterraneo, dalle Baleari alla Grecia. I nuraghe costruiti tra il 1500 ed il 400 a.C., erano 6000/7000, distribuiti su tutta l'isola in modo che da ciascuno se ne potesse vedere un altro. Se i nuraghi furono abitati tutti nello stesso periodo, la popolazione potrebbe essere stata di oltre 200.000 individui.

Intorno all'800 a.C. i Fenici colonizzarono le coste dell'isola, soprattutto quelle meridionali; successivamente la colonizzazione continuò a opera di Cartagine, una colonia fenicia in Tunisia. In seguito al predominio romano nel Mediterraneo e alla caduta di Cartagine, la Sardegna venne occupata dai Romani, che però non riuscirono a conquistare pienamente le zone più interne. Alla caduta dell'Impero romano seguirono le invasioni di Vandali, Bizantini e Saraceni; la dominazione araba continuò fino al X secolo. L'isola fu poi assoggettata al controllo di Pisa, seguito da quello congiunto di Catalogna e Aragona, e infine al Regno sabaudo. Le ultime tre dominazioni

hanno lasciato colonie in zone limitate dell'isola, in cui sono rimaste tracce linguistiche: un'area di lingua catalana nel nordovest; una di coloni piemontesi e liguri, nel sudovest, i cui discendenti parlano ancora oggi quei dialetti; una di toscani nel nordest. Tutti gli invasori degli ultimi due millenni hanno lasciato tracce genetiche limitate.

La causa di gran lunga più importante della considerevole differenza genetica tra i Sardi e le altre popolazioni europee è costituita dalla deriva genetica, in quanto numerosi geni presentano frequenze molto differenti dalle medie di diverse regioni europee (o africane).

In Sardegna si ha la frequenza più bassa del *gene RH-negativo* rispetto alle regioni del Mediterraneo, tra le più alte frequenze nel mondo del gene M del sistema MNS. Anche il gene diaforasi-2(DIA2) ed i *sistemi HLA* e *GM* presentano frequenze anomale in Sardegna: la frequenza più alta nel mondo *dell'HLA-18*, frequenze alte di alcuni alleli caratteristici delle regioni africane, sebbene la loro distribuzione generale sia diversa da quella che si riscontra negli alleli tipici delle popolazioni africane[2]. Anche le varianti della talassemia hanno frequenze peculiari: una variante molecolare (B^{39}) ha la frequenza più alta in Sardegna, mentre è rara altrove. Da uno studio di Piazza e collaboratori[3] è emerso che, fra le popolazioni del Mediterraneo, i Sardi sono più simili agli Italiani,ai Libanesi e agli africani settentrionali (i Greci non erano stati presi in considerazione). In base ai dati usati nello studio di Cavalli Sforza-Menozzi-Piazza[4]le distanze genetiche dei Sardi dalle popolazioni di maggior rilievo sono:

Popolazione	**Distanza**	**Errore Standard**
GRECI	190	30
ITALIANA	221	54
BASCHI	261	68
LIBANESI	340	66
AFRICANI SETTENTRIONALI	**732**	**168**

Dai dati si deduce che soltanto i nordafricani non hanno contribuito in modo rilevante al pool genetico sardo, mentre l'Italia e la Grecia sono state probabilmente luoghi di origine dei primi occupanti del Neolitico. Gli uomini neolitici provenivano a loro volta dal Medio Oriente e dalla Turchia, ma probabilmente durante il passaggio attraverso la Grecia e l'Italia meridionale il loro genotipo fu diluito dal flusso genico delle popolazioni mesolitiche locali.
I moderni Libanesi sono i più diretti discendenti dei Fenici, che fornirono un contributo al patrimonio genetico dei Sardi. Pertanto l'antica origine genetica può essere ancora individuata, nonostante l'effetto rilevante della deriva genetica che deve essersi verificata in tempi precedenti.
In Sardegna l'effetto della deriva genetica può essere stimato in modo approssimativo considerando che, in condizioni di saturazione, la popolazione sarda del tardo Paleolitico potrebbe essere stata formata da 700-1800 individui[4]. Nel corso dei millenni questa ridotta dimensione della popolazione avrebbe generato una deriva genetica considerevole, senza dover invocare l'effetto di un numero piccolo di fondatori (comunque non da escludere del tutto). Gli uomini del Neolitico devono avere apportato nuovi geni, ma data la distanza dell'isola dalle coste mediterranee, probabilmente erano anch'essi poco numerosi e alla deriva si è quindi aggiunto anche l'effetto del fondatore. Non si sa molto del primo Neolitico in Sardegna: lo scenario demografico è più chiaro solo in relazione alla popolazione successiva, quella nuragica, la cui densità fu sufficiente ad arrestare l'effetto della deriva e a mantenere le frequenze geniche all'incirca a livelli attuali. Gli insediamenti successivi furono limitati alle zone costiere, fatta eccezione per Fenici e Cartaginesi,numerosi in tutto il meridione, in particolare nella zona sudoccidentale dell'isola. La Sardegna è stata anche oggetto di studi sulla differenza linguistica e sulla distribuzione dei cognomi. Nell'isola vengono parlati numerosi dialetti.
A causa dell'isolamento geografico della Sardegna nel Mar Mediterraneo, la popolazione Sarda, quindi può essere considerata una genetica isolata. La fauna e la flora endemica sottolineano questa peculiarità, riscontrabile anche

nella struttura genetica e culturale della popolazione umana. Per queste ragioni i Sardi sono stati oggetto di numerosi studi in campo antropologico e di genetica di popolazioni[5,6,7,8].

Diversi studi hanno mostrato che il genoma degli attuali abitanti della Sardegna contiene ancora i segni di una lunga storia di isolamento. Queste caratteristiche la rendono, come genetica isolata, una popolazione ideale per gli studi di associazione[9,10,11]. Tuttavia rimane ancora molto da scoprire riguardo le regioni ereditate da antenati comuni, come le short Runs of Homozygosity (RoHs), o le porzioni di genoma che sono state selezionate positivamente. Le RoHs sono regioni del genoma in cui le copie ereditate da entrambi i genitori sono identiche, in quanto entrambi i genitori le hanno ereditate da un antenato comune in un certo momento nel passato.

Nel presente studio abbiamo analizzato la struttura genetica della popolazione sarda usando 1,2 milioni di SNPs da 1077 sardi precedentemente inclusi in uno studio di associazione genome-wide, e 79 individui sani provenienti dalla penisola Italiana. Gli obiettivi da perseguire nello studio erano:

(1) Confermare attraverso l'uso dei dati genome-wide l'omogeneità della popolazione Sarda a livello interregionale.

(2) Dedurre, attraverso l'uso delle RoHs, la storia genetica della popolazione stimando il livello di background dell'antenato comune all'interno dell' isola e confrontarla con la penisola italiana.

(3) Identificare segnali di selezione positiva.

Materiali e Metodi

I dati genotipici di 1077 soggetti sani provenienti dalla Sardegna sono stati usati come data-set primario. Questi campioni sono stati precedentemente raccolti in un consorzio internazionale per uno studio di associazione genome-wide(GWAS) sull'ipertensione (HyperGene) [12].

I pazienti sono stati raggruppati in base al luogo di nascita, dividendo la Sardegna in base al dialetto parlato come suggerito da Contini e collaboratori[13,14]. Nel presente lavoro è stata usata una semplificazione di questo approccio dividendo l'isola in sei macro-aree principali come mostrato nella *figura 1*: Gallurese(=n77), Nuorese(n=88), Logudorese(n=385), Sassarese(n=342), Alghero(n=87), Campidanese (n=98).

Parte dei campioni (n=250) sono stati già analizzati in un precedente lavoro[15]. Un gruppo addizionale consiste di 79 individui Italiani (Italia peninsulare) che sono stati inclusi nello studio allo scopo di eseguire una comparazione tra il background genetico della Sardegna e il continente Italiano. I soggetti della penisola Italiana sono stati genotipizzati presso i laboratori dei nostri collaboratori di Torino per oltre 1 Milione di SNPs(HumanOmnil-Quad1.0, BeadChip, Illumina Inc, S.Diego, CA, USA), mentre la coorte sarda presso i laboratori della Fondazione Filarete, Univesrità degli Studi di Milano per mezzo della piattaforma ILLUMINA. Per comparare la Sardegna e l'Italia sono stati presi in considerazione solo gli SNPs comuni ad entrambi(circa 520 markers). Tutti i campioni sono stati raccolti previo consenso informato e analizzati in maniera anonima. Il loro utilizzo per gli studi di genetica di popolazione è stato approvato dal comitato etico della Human Genetics Foundation (HuGeF) a Torino .

Procedura di controllo di qualità

Sono state applicate stringenti procedure del controllo di qualità durante l'esecuzione di analisi di genotipizzazione degli SNPs. Gli SNPs con l'allele a minor frequenza (MAF) inferiore a 0,01 (1%) sono stati esclusi, come anche quelli che non superavano il test dell'equilibrio di Hardy-Weinberg ($p<1x10^{-3}$). Allo scopo di stimare il numero specifico di RoHs, sono stati esclusi gli SNPs presenti sui cromosomi sessuali. Dopo le procedure del controllo di qualità, il data-set della Sardegna conteneva un totale di 976.970 SNPs.

Analisi dati statistici

Le analisi sono state eseguite su differenti livelli. Il primo aveva lo scopo di valutare la struttura genetica interna della Sardegna. Un secondo livello aveva l'obiettivo di ricostruire la storia genetica della popolazione attraverso l'analisi delle Rohs, e l'identificazione di regioni genetiche sotto selezione positiva.

Struttura della Popolazione Sarda

E' stata eseguita l'analisi delle componenti principali (PCA) usando tutto il set di markers genetici, attraverso l'algoritmo sviluppato all'interno del pacchetto del software R[16], SNPRelate[17].
Le PCA di ogni individuo campione sono state messe su un grafico definito dai primi due autovettori: i soggetti della stessa macro-area linguistica o la stessa area geografica sono visualizzati con colore identico(*Figure 2A e B*).
Abbiamo usato le prime 4 componenti principali (PCA) come predittori in una regressione logistica multinomiale, usando la macro-area linguistica come variabile dipendente(Figura S1**)**. Abbiamo poi valutato l'accuratezza del modello descritto: per ogni campione confrontando la macro-area più probabile stimata dal modello con quella reale (10.000 iterazioni).

E' stato calcolato il Pairwise inflation factors (lGC) [18] attraverso il software PLINK[19](–*adjust* option), simulando uno studio caso-controllo tra ogni coppia di macro-aree. Abbiamo utilizzato due differenti metodi per calcolare l'indice di fissazione F_{st}(distanza genetica): il primo aveva l'intento di produrre una stima sui dati con inbreeding significativo mentre la seconda è stata pensata per una popolazione panmittica (Sardegna/Italia peninsulare). Pairwise genetic Fst corretto per l'inbreeding tra le sei macro-aree, è stato stimato come suggerito da Rich *et. al.* [20]. F_{st} tra la Sardegna e l'Italia peninsulare è stata calcolata usando lo stimatore di Hudson per dati genome-wide[21], come suggerito da Bhatia *et. al.* [22].

Il significato del coefficiente di inbreeding è stato stimato sulla base del numero di genotipi omozigoti osservati rispetto agli attesi sull'intero genoma, utilizzando il set di dati contenente anche gli individui dell'Italia peninsulare [PLINK software (–het option)].

Le differenze tra la popolazione sarda e quella della penisola italiana sono state valutate usando un T test.

Il software ADMIXTURE[23] è stato utilizzato per stimare la discendenza di ogni individuo nella popolazione sarda e nei soggetti dell'Italia peninsulare. E' stato applicato il metodo della cross-validazione basato sull'errore, per trovare il numero di cluster (k) dopo 20 runs.

Analisi delle "Runs of Homozygosity"

Le RoHs sono state stimate separatamente per la Sardegna e la penisola italiana [PLINK software (–homozyg option)]. I seguenti parametri sono stati utilizzati per l'algoritmo di stima:

1) una "finestra scorrevole" di 5000kb, con un minimo di 50 SNPs che può essere presente nelle regioni considerate;

2) per una determinata finestra, un massimo di un eterozigote e un massimo di cinque chiamate mancanti consentite;

3) ogni SNPs è stato considerato parte di un segmento di omozigosi quando la percentuale di "finestre di omozigosi" sovrapposte era oltre il valore di soglia di 0,05. Abbiamo identificato 6 categorie di RoH basato sulla lunghezza delle regioni genomiche di omozigosità (0.5–1 Mb,1–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, .16 Mb), e abbiamo stimato la proporzione di individui con RoHs di differenti dimensioni in ogni macro-area della Sardegna.

Le differenze tra le macro-aree della Sardegna e l'Italia peninsulare sono state valutate usando un T test.

Abbiamo inoltre stimato la percentuale di genoma coperto da regioni di omozigosità (FroH%) secondo McQuillian et al. [24]. Due classi di RoHs sono state considerate in questa analisi: RoHs 0.5 Mb e RoHs 5 Mb. Per ogni classe e per ogni macro-area abbiamo calcolato la FRoH% media su tutti gli individui, così come l'importo medio della lunghezza di tutte le RoHs della stessa classe.

E' stato eseguito un T test per valutare le differenze tra le due classi di RoH all'interno di ogni macro-area e l'Italia.

Omozigosità da aplotipo esteso

Il software FastPHASE [25] è stato usato per eseguire una fase di stima dell' aplotipo. Gli aplotipi stimati sono stati successivamente utilizzati per rilevare impronte di selezione da struttura aplotipo.

Per ogni SNPs abbiamo calcolato sia le Extended haplotype homozygosity(EHH) statistic [26] di entrambi gli alleli (ancestrale e derivato), sia il punteggio di aplotipo integrato (integrated haplotype score, IHS) [27](La tecnica dell'iHs score calcola il rapporto tra allele derivato e allele ancestrale). L'algoritmo è sviluppato all'interno del pacchetto R rehh [28]. Per queste specifiche analisi abbiamo impiegato un totale di circa 900 k markers per i quali erano disponibili informazioni riguardo l'allele ancestrale, in un database pubblico[29]. Infine abbiamo individuato le regioni cromosomiali che mostravano un arricchimento di SNPs con |iHS|>4, usando l'approccio suggerito da Voight et al. [27]. È stata applicata la correzione per confronti multipli basata sulla permutazione

Risultati

Il modello fatto con la regressione logistica multinomiale, usando i primi 4 autovettori come predittori delle macro-aree linguistiche, ha mostrato un' accuratezza molto bassa (da un minimo di 0,2044 a un massimo di 0.3201, 10,000), suggerendo un alto grado di omogeneità all'interno della popolazione sarda. Nessuna sub-popolazione è stata apparentemente identificata proiettando i campioni sardi sul grafico (costruito con i primi due autovettori) utilizzando tutti i marcatori autosomici (934.288 SNPs) all'interno delle 6 macro-aree linguistiche (*Figura 2A*), o dividendo l'isola in 3 regioni geografiche (*Figura 2B*). La distribuzione dei primi 4 autovettori è mostrata in *figura S1*. Il valore di inbreeding per tutte le parwaise Fst corretto entro le macro-aree linguistiche, era vicino allo zero (*Tabella 1*), e abbiamo osservato uno stimatore Fst di 0,003 (p-value, 0,0001 95% CI 0,0025-0,0033) quando si confrontano la Sardegna e l'Italia peninsulare. L'inflation factor (lGC) erano molto vicini ad 1(da 1.01 a 1.05) (*Tabella1*).

L'analisi della discendenza ha sottolineato la presenza di un background genetico comune per tutti gli individui dell'isola(*Figura 3*).La comune discendenza osservata ha reso inattuabile qualsiasi tentativo di raggruppare gli individui sulla base del luogo di nascita. Attraverso l'utilizzo del cross validation error, abbiamo indicato k=2 come numero di cluster compatibili con i dati; inoltre, valori più elevati di K non hanno rivelato ascendenze aggiuntive popolazione-specifici. La percentuale di genoma coperto da RoHs>0,5Mb ($F_{RoH}\%_{0,5}$) era più alto nella Sardegna quando paragonato con la penisola italiana, con la sola eccezione dell'area circondante la zona di Alghero (*Tabella 2*).

Non vi erano significative differenze tra i sardi e gli italiani quando venivano confrontate le frazioni del genoma coperto da RoH>0.5 Mb (FRoH%5) (*Tabella 2*). Sono state trovate significative differenze nei coefficienti medi di inbreeding tra le macro-aree (Campidanese, Gallurese, Sassarese, e Logudorese) e l'Italia peninsulare (*Tabella 3*). Dal momento che la

distribuzione delle diverse classi di RoH permette di studiare diversi modelli demografici che coinvolgono una popolazione, abbiamo ulteriormente diviso RoHs in 6 diverse classi (0.5-1 Mb,1-2 Mb, 2-4 Mb, 4-8 Mb, 8-16 Mb, >16 Mb), come illustrato nella Tabella 4.

La Sardegna ha un numero più alto di RoHs rispetto all'Italia per 2 delle 6 classi di RoH: 0.51 Mb, e 1–2 Mb (p-value,0.05). Osservando le classi di RoH più lunghi (8–16 Mb e >16 Mb) non sono state osservate significative differenze tra le due regioni, ad eccezione del Campidanese per le classi 8–16 Mb. Confrontando le classi di RoH maggiori di 2Mb, il distretto di Alghero e il Sassarese non hanno presentato differenze con l'Italia peninsulare.

Per rilevare eventuali impronte di selezione positiva, è stato stimato il decadimento EHH standardizzato (ossia, iHS [27]).

Nove regioni genomiche, situate in più di 200 differenti geni, mostrano segni di selezione positiva(*Tabella 5*). Sono state descritte qui per la prima volta regioni genomiche sotto selezione positiva sul cromosoma 9(da 70,303,655 a 70,400,714 bp) e sul cromosoma 19(da 22,561,972 a 22,586,080 e da 32,961,206 a 33,175,723). Non inaspettatamente, in una grande regione cromosomica sul cromosoma 6 (6p21.3) vi sono prove di selezione positiva, che comprende il sistema dell' antigene leucocitario umano (HLA).

Un'altra regione interessante è 11q12.1 che contiene 24 geni correlati con l'attività del recettore olfattivo.

Discussione

Un gran numero di markers genetici appartenenti a differenti categorie sono stati utilizzati per descrivere la peculiarità genetica della Sardegna in confronto con le altre popolazioni Mediterranee ed Europee: markers della genetica classica [6,30,31,32,33], sistema HLA[34,35,36] markers autosomici[5,9,10,37,38], distribuzione di mutazioni rare della fibrosi cistica[39], polimorfismi del DNA mitocondriale (mtDNA)[40,41,42,43,44,45], varianti genetiche del cromosoma Y e dati di sequenze[11,46,47,48,49,50,51].

In generale la Sardegna appare caratterizzata da una larga omogeneità interna [9,11], come tutte le popolazioni isolate, nonostante altri ricercatori suggeriscano la presenza nell'isola di sotto-popolazioni geneticamente differenti[10, 52]. Recentemente, diversi studi genome-wide sono stati effettuati sulla popolazione sarda avvalendosi della omogeneità genetica dell 'isola, utilizzando anche un'ampia coorte di individui [53,54,55,56,57].

In questo studio, noi abbiamo riconfermato l'alta omogeneità interna della Sardegna usando 4 differenti metodi:

PCA = Analisi delle Componenti Principali
Fst distanza = Distanza genica
Inflation factor parameter (IGC) = Indicatore di multicollinearità
Ancestry estimation = Stima della discendenza

La mancanza di una struttura di sotto-popolazioni appare chiara dalla PCA. Infatti, il modello di regressione logistica multinomiale, mostra che le prime 4 componenti principali non sono in grado di prevedere le macro-aree linguistiche. Inoltre la correzione del valore di inbreeding (Fst) era tra $9,1*10^{-5}$ a $1,1*10^{-4}$ e il IGC era di circa 1, entrambi indicavano la mancanza di differenziazione tra le popolazioni delle diverse aree, e di stratificazione genetica all'interno dell'isola.

L'ancestry estimation indicava un notevole grado di similarità per tutti i soggetti sardi raccolti, allo stesso tempo indicava una eterogeneità significativa quando sardi sono stati confrontati con soggetti dell'Italia peninsulare.

È tuttavia degno di nota che alcune sub-regioni Sardegna, come Ogliastra, sono in realtà formate da villaggi isolati, ognuno dei quali con una demografia unica. Diversi studi [10,52,58] hanno osservato differenze di linkage disequilibrium (LD) e la struttura della popolazione tra questi villaggi. Purtroppo, il numero limitato di individui di Ogliastra nel nostro campione (N = 16) non ci ha permesso di testare l'ipotesi di sotto-strutture genetiche a livello micro-geografico.

L'isolamento della popolazione ha lasciato il segno sul DNA dei sardi. In realtà un aumento di 2 volte nella media dell'omozigosi rispetto all'Italia, è ancora rilevabile. Tuttavia abbiamo ancora trovato evidenze di una significativa diminuzione di omozigosità nei dintorni di Alghero, che è la macro-area linguistica con i segni più bassi di isolamento in Sardegna.

Ci siamo concentrati sulle RoHs per uno studio più approfondito sulla storia demografica dell'isola. Le RoHs si osservano nel genoma di ogni individuo, e la loro lunghezza è in relazione con il loro momento di origine. Le RoHs descrivono differenti aspetti di una popolazione, come la consanguineità, l'endogamia, ed eventi demografici come i colli di bottiglia. Abbiamo quindi valutato la percentuale media del genoma coperto da RoH>0,5 Mb e RoH.>5 Mb ($F_{RoH0.5\%}$ e $F_{RoH5\%}$, rispettivamente) all'interno di ogni sotto-popolazione confrontata con quelle dell'Italia peninsulare. Il $F_{RoH0.5\%}$ descrive la tendenza globale di omozigosi all'interno delle sub-popolazioni, mentre il $F_{RoH5\%}$ fornisce informazioni su altri fenomeni, come l'endogamia o recente consanguineità. La media FRoH0.5% e la media della somma della lunghezza di questi segmenti in Sardegna erano più elevati rispetto a l'Italia (la media della somma di FRoH0.5 per la Sardegna da 72,77 a 82,55 Mb, per l'Italia 67.55 Mb). Queste osservazioni sono coerenti con una popolazione ancestrale effettiva di piccole dimensioni (N_e) in Sardegna e un livello più

profondo di discendenza comune. Ancora una volta la zona di Alghero contrasta con tali osservazioni mostrando un FRoH0.5% simile all'Italia peninsulare. Tuttavia, non siamo stati in grado di osservare un simile trend per FRoH5% in ognuna delle 6 macro-aree.

Per ottenere maggiori dettagli, abbiamo classificato le RoHs in sei classi diverse. In media, in Sardegna, la media delle somme delle RoHs più corte(0,5-1 Mb e 1-2 Mb) era significativamente più lunga che in Italia.

Questo fenomeno può essere spiegato come il risultato di aplotipi estesi comuni, ereditati probabilmente da entrambi i genitori, frequenti in comunità piccole e isolate [59].

Altre macro-aree come Campidanese e Gallurese (RoH from 2 to 8 Mb) e Logudorese e Nuorese (RoH 2–4 Mb) presentano ancora tracce di endogamia messe a confronto con l'Italia peninsulare.

Al contrario nell'area di Alghero, le RoHs oltre la soglia dei 2Mb, erano più corte e meno comuni che nelle altre popolazioni sarde; questo risultato indica gradi di consanguineità ed endogamia più bassi in questa sub-popolazione. Va notato che la città nord-occidentale di Alghero è una comunità di lingua catalana e questa lingua è una notevole eccezione rispetto a tutte le varietà di dialetti sardi. Il dialetto di Alghero deriva da eventi storici che hanno interessato la città nel Medioevo, quando la popolazione è stata inondata dall'arrivo di coloni di lingua catalana [60]. A nostra conoscenza, solo uno studio ha precedentemente valutato modelli di omozigosità in tutto il genoma nella popolazione sarda [9]. Anche se i criteri utilizzati per l'individuazione di RoHs sono leggermente diverse tra il presente studio e quello di Pardo e colleghi, i risultati dei due studi sono coerenti. In più abbiamo cercato impronte di selezione positiva nel genoma dei sardi attraverso l'utilizzo delle EHH e il test iHS.

I nostri risultati hanno identificato alcune regioni genomiche non precedentemente descritte come sotto selezione positiva che possono essere considerate come nuovi candidati meritevoli di indagine per la selezione positiva nella popolazione sarda. Tra loro TMEM252 gene (ID

169693) e il gene PGM5 (ID 5239), ed una regione sul cromosoma 19 contenente un luno RNA non codificante(LINC00662). Come previsto, abbiamo catturato molti dei segnali precedentemente descritti di recente selezione positiva. Specificamente il gene PRLH (ID 51052) e il gene MLPH (ID 79083), entrambi localizzati sul braccio lungo del cromosoma 2, che sono sotto selezione positiva nel Medio Oriente e nelle popolazioni Europee[61], il gene SH3BP5L(ID 80851) [62],e una regione sul cromosoma 11 contenente diversi geni correlati all'olfatto[59]. Come riportato in letteratura, la regione del sistema HLA(human leukocyte antigen) è sotto selezione positiva nelle popolazioni Europee, Medio Orientali e Sud Asia[61]. Nel nostro studio non abbiamo trovato il gene della lattasi (LCT ID 3938) tra le regioni sotto selezione positiva, come riportato anche in altri studi[61,64].

Conclusioni

Anche se il principale limite del nostro studio è che le informazioni sulle origini degli individui sardi erano basate solo sul loro luogo di nascita, il nostro studio ha riconfermato, utilizzando diversi approcci, l'elevato grado di omogeneità genetica interna in Sardegna. Abbiamo dimostrato che il genoma dei sardi ha una media di coefficiente di inbreeding superiore a quelli della penisola italiana. Inoltre, il genoma della Sardegna conserva ancora tracce della complessa storia demografica dell'isola. Tra le macro-aree analizzate, la zona circostante Alghero mostra meno consanguineità di altri, in accordo con la sua peculiare storia sottolineata anche dal lingua locale usata attualmente in seconda analisi rispetto all'Italiano e che in realtà è a tutti gli effetti "Catalano antico". Sono state individuate diverse regioni genomiche che mostrano segnali di selezione positiva, alcuni di loro non precedentemente descritte e come tale meritevoli di ulteriori indagini. L'omogeneità genetica della Sardegna è una ottima premessa per la definizione di determinanti genetiche di "malattie complesse. Nel prossimo futuro, i nostri risultati potrebbero essere confermati dal ri-sequenziamento dei geni/regioni che mostrano segni di selezione positiva e individuando potenziali SNP/aplotipi funzionali.

Figure e tabelle

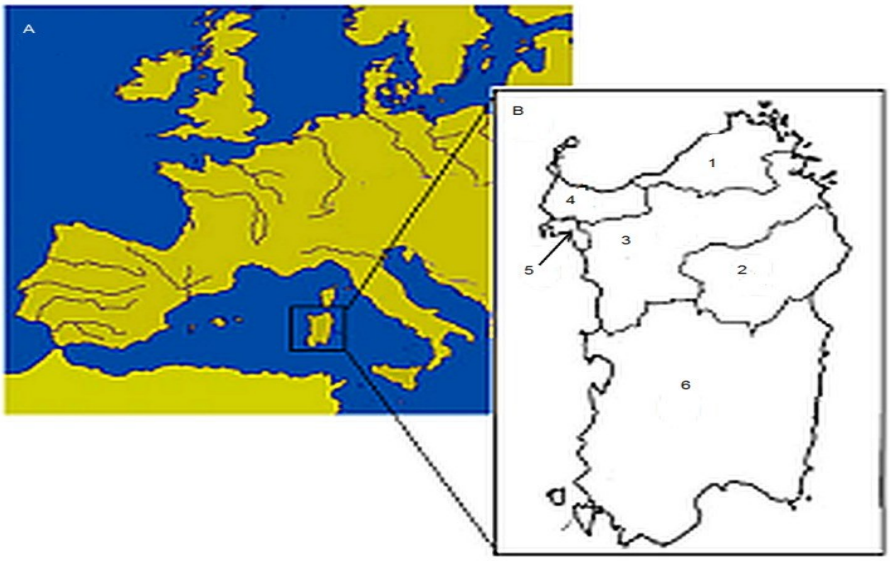

Figura 1. Map of Mediterranean basin showing the localization of Sardinia and Sardinian linguistic domains. **A)** Map of the Mediterranean basin showing the geographic position of Sardinia. **B)** The Sardinian linguistic domains: 1 = Gallurese (77 individuals); 2 = Nuorese (88);

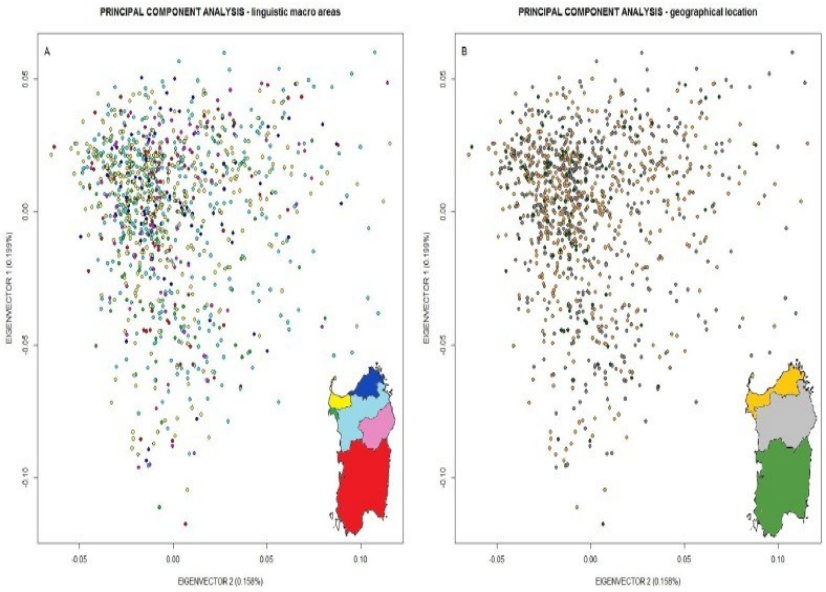

Figure 2. SNP-Based Principal Component Analysis of 1,077 individuals from Sardinia.

Figure **2 A)** division accounting linguistic macro-areas. Key of the colors: red: Campidanese; green: Alghero; deep blue: Gallurese; light blue: Logudorese; yellow: Sassarese; purple: Nuorese.

Figure **2 B)** division accounting geographical areas. Key of the colors: green: Southern Sardinia; grey: Central Sardinia; yellow: Northern Sardinia.

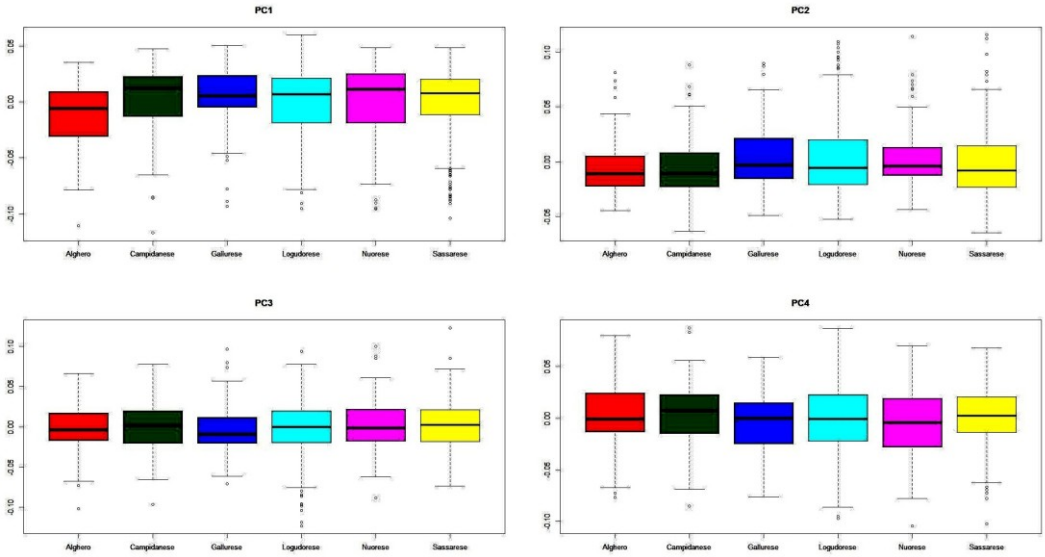

Figure S1. Box plot distribution of the first four eigenvectors in the 6 macro-areas.

Tabella 1. Fst values (in bold) and genomic control inflation factor (IGC) (in italics) between Sardinian linguistic macro-areas.

λ_{GC}/F_{st}	Campidanese	Alghero	Gallurese	Logudorese	Nuorese	Sassarese
Campidanese	-	**4.4×10^{-5}**	**1.1×10^{-4}**	**3.2×10^{-5}**	**2.2×10^{-5}**	**8.5×10^{-6}**
Alghero	*1.037*	-	**1.5×10^{-4}**	**9.1×10^{-5}**	**1.1×10^{-4}**	**7.1×10^{-5}**
Gallurese	*1.051*	*1.047*	-	**1.2×10^{-4}**	**1.4×10^{-4}**	**1.1×10^{-4}**
Logudorese	*1.018*	*1.041*	*1.040*	-	**6.1×10^{-5}**	**4.2×10^{-5}**
Nuorese	*1.019*	*1.027*	*1.040*	*1.028*	-	**4.0×10^{-5}**
Sassarese	*1.012*	*1.032*	*1.046*	*1.046*	*1.021*	-

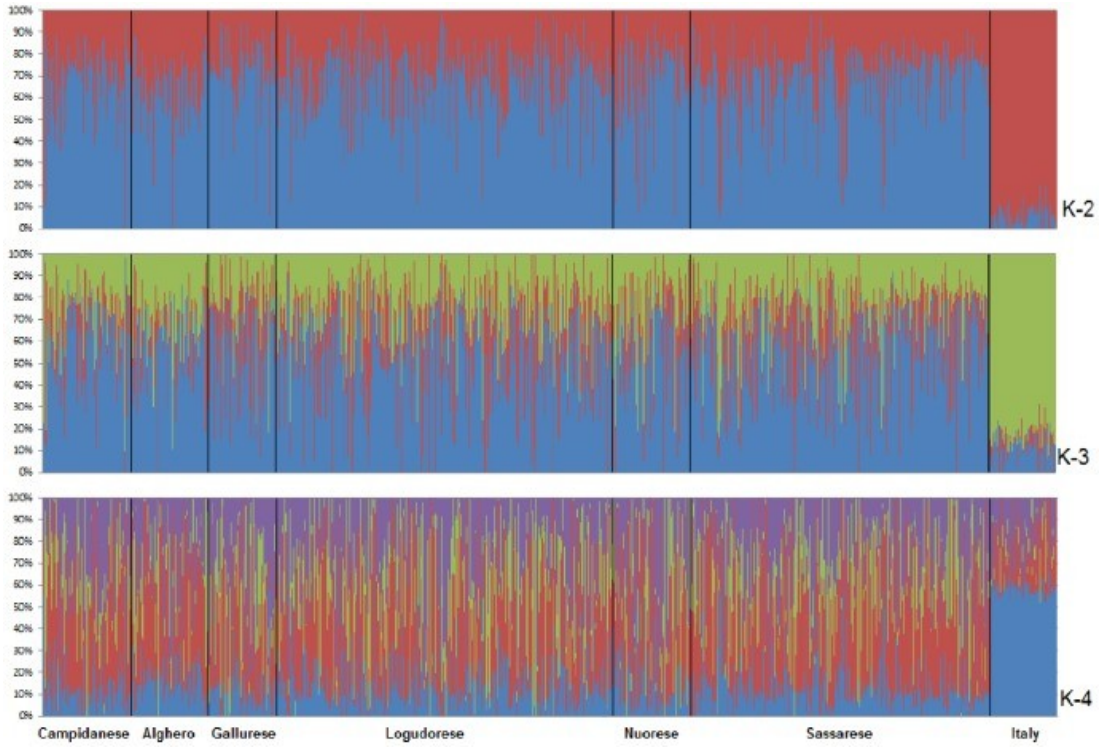

Figure 3. ADMIXTURE software results for K = 2–4. Ancestry for each individual inferred using ADMIXTURE software.

Table 2. Mean genomic inbreeding coefficients (FRoH %) using 0.5 and 5 Mb minimum RoH thresholds and mean sum of RoH

	F_{RoH}%≥0.5	F_{RoH}%≥5	Mean (SD) sum of RoH (Mb)			
			mean≥0.5		mean≥5	
Campidanese	3.08*	0.49	82.55*	3.81	13.24	3.14
Alghero	2.71	0.29	72.77	2.86	7.86	2.28
Gallurese	3.10*	0.51	83.09*	4.39	13.65	3.82
Logudorese	2.96*	0.41	79.33*	1.79	11.06	1.53
Nuorese	2.89*	0.42	77.61*	3.74	11.15	3.26
Sassarese	2.94*	0.44	78.84*	2.05	11.67	1.84
Italy	2.52	0.47	67.55	4.98	12.64	4.28

* *p-value* smaller than 0.05 when comparing each linguistic macro-area to peninsular Italy.

Table 3. Mean inbreeding coefficients.

	Mean inbreeding coefficient	SE	*p-Value*
Campidanese	0.0106	0.00015	0.002
Alghero	0.0058	0.00014	0.26
Gallurese	0.0100	0.00022	0.01
Logudorese	0.0086	0.00002	0.003
Nuorese	0.0079	0.00016	0.06
Sassarese	0.0082	0.00004	0.01
Italy	0.0046	0.00001	-

Mean inbreeding coefficients, standard errors (SE) and T test *p-Values* of Sardinia macro-areas and peninsular Italy.

Table 4. Percentage of the accessible genome occupied (2.84 Gb) and mean sum of RoH in Mb (with standard errors SE) for six classes of RoH.

	0.5-1 Mb			1-2 Mb			2-4 Mb			4-8 Mb			8-16 Mb			>16 Mb		
	% RoH	mean	SE	% RoH	mean	SE	% RoH	mean	SE	% RoH	mean	SE	% RoH	mean	SE	% RoH	mean	SE
Campidanese	1.23	32.92*	0.53	0.98	26.25*	0.56	0.31	8.27*°	0.6	0.2	5.28*°	0.78	0.17	4.66*	1.26	0.19	5.18	1.74
Alghero	1.19	31.86*	0.56	0.95	25.61*	0.58	0.26	6.89	0.51	0.1	2.66	0.57	0.13	3.46	1.06	0.09	2.29	1.12
Gallurese	1.19	31.83*	0.55	1.01	27.12*°	0.65	0.32	8.68*°	0.64	0.22	5.98*°	1.06	0.15	4.07	1.07	0.2	5.42	2.26
Logudorese	1.21	32.42*	0.27	1	26.91*°	0.33	0.29	7.68*°	0.27	0.14	3.82°	0.34	0.16	4.17	0.58	0.16	4.33	0.93
Nuorese	1.19	31.82*	0.58	0.93	25.02*	0.59	0.31	8.33*	0.49	0.15	3.89	0.74	0.19	5.01	1.25	0.13	3.54	1.82
Sassarese	1.21	32.52*	0.28	0.99	26.47*	0.32	0.26	6.96	0.27	0.14	3.7	0.39	0.14	3.72	0.59	0.2	5.48°	1.18
Italy	0.98	26.25°	0.43	0.8	21.49°	0.63	0.24	6.42	0.63	0.13	3.45	1.01	0.18	4.9	1.51	0.16	4.4	2.07

*T test p-value<0.05 comparing to Italy,
°T test p-value<0.05 comparing to Alghero.

Position NCBI36/ hg18	SIZE	n SNP |iHS|>4	n SNP	MAX|iHS|	MAX iHS SNP	*p-value*	empirical *p-value*	genes
chr19: 32,961,206–33,175,723	215	12	69	6.25	rs17714275	3.87 e^{-34}	<0.0001	*LINC00662*
chr6: 29,555,703–33,009,633	3454	35	4884	5.37	rs397081	1.64 e^{-21}	<0.0001	*GABRB1;MOG; HLA-F;HLA-G; etc;*
chr2: 238,113,451–238,164,950	51	7	37	−5.33	rs2292871	6.62 e^{-22}	<0.0001	*MLPH;PRLH;RAB17*
chr9: 70,303,655–70,400,714	97	8	37	5.06	rs11143002	4.86 e^{-25}	<0.0001	*PGM5; TMEM252*
chr19: 22,561,972–22,586,080	24	8	16	−4.66	rs4932781	4.49 e^{-29}	<0.0001	*LOC440518;LOC100996349*
chr5: 109,659,513–109,731,650	72	13	28	−4.65	rs10478008	9.03 e^{-44}	<0.0001	*NA*
chr1: 247,047,666–247,088,866	41	7	15	−4.48	rs12058711	1.05 e^{-25}	<0.0001	*SH3BP5L*
chr4: 34,062,734–34,244,104	181	13	46	−4.39	rs11936559	5.37 e^{-40}	<0.0001	*NA*
chr11: 55,732,908–56,414,929	682	10	228	4.27	rs12576240	4.79 e^{-25}	<0.0001	*OR5T2;OR5T3;OR5T1;OR8H1; etc;*

Column headers: Position on NCBI36/hg18 of region showing evidence for selection; Size in Kb of the genomic region; nSNP |iHS|>4 indicates the number of SNPs with an absolute |iHS| higher than 4 in each region; nSNP is the number of SNPs in each region; Max iHS is the highest value of each region; Max iHS SNP is the polymorphism with the highest value for each region; P-values: nominal p-values; Empirical p-values: after permutation-based multiple testing corrections; Genes: the genes within the region. When, in the genomic region, there are more than 4 genes, only the first 4 are indicated.

Table 5. Nine genomic regions showing signals of positive selection in the Sardinian's genome ordered by |iHS|

References

1. C.F. Spoor, P.Y. Sondaar (1986),Human fossils from the endemic island fauna of Sardinia,in" J.Hum. Evol.", 15, pp.399-408.

2. Piazza A, Mayr WR, Contu L, Amoroso A, Borelli I, Curtoni ES, Marcello C, Moroni A, Olivetti E, Richiardi P, et al. Genetic and population structure of four Sardinian villages. Ann Hum Genet. 1985 Jan;49(Pt 1):47-63.

3. Piazza A, Sgaramella-Zonta L, Gluckman P, Cavalli-Sforza LL.The Fifth Histocompatibility Workshop gene frequency data: a phylogenetic analysis. Tissue Antigens. 1975 Jun;5(5):445-63.

4. Cavalli Sforza-Menozzi-Piazza(2009).The History and Geography of Human Genes. Gli Adelphi 5.6.b *Sardegna*

5. Calo CM, Autuori L, Di Gaetano C, Latini V, Mameli GE, et al. (1998) The polymorphism of the APOB 39 VNTR in the populations of the three largest islands of the western Mediterranean. Anthropologischer Anzeiger; Bericht ùber die biologisch-anthropologische Literatur 56: 227–238.

6. 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. Annals of Human Genetics 60: 125–141.

7. Caramelli D, Vernesi C, Sanna S, Sampietro L, Lari M, et al. (2007) Genetic variation in prehistoric Sardinia. Human Genetics 122: 327–336.

8. D'Amore G, Di Marco S, Floris G, Pacciani E, Sanna E (2010) Craniofacial morphometric variation and the biological history of the peopling of Sardinia. Homo 61: 385–412.

9. Pardo LM, Piras G, Asproni R, van der Gaag KJ, Gabbas A, et al. (2012) Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers. Eur J Hum Genet 20: 956–964.

10. Piras IS, De Montis A, Calo CM, Marini M, Atzori M, et al. (2012) Genome-wide scan with nearly 700 000 SNPs in two Sardinian sub-

populations suggests some regions as candidate targets for positive selection. Eur J Hum Genet.

11. Contu D, Morelli L, Santoni F, Foster JW, Francalacci P, et al. (2008) Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: Inference for association scans. PLoS One 3.

12. Salvi E, Kutalik Z, Glorioso N, Benaglio P, Frau F, et al. (2011) Genomewide Association Study Using a High-Density Single Nucleotide Polymorphism Array and Case-Control Design Identifies a Novel Essential Hypertension Suscepti- bility Locus in the Promoter Region of Endothelial NO Synthase. Hypertension.

13.Contini M (1979) Classification phonologique des langages sardes. Bull Inst Phonetique Grenoble 8: 57–96.

14. Contini M, Cappello N, Griffo R, Rendine S, Piazza A (1989) Ge olinguistique et geogenetique: Une demarche interdisciplinaire. Geolinguistique 4: 129–197.

15. Di Gaetano C, Voglino F, Guarrera S, Fiorito G, Rosa F, et al. (2012) An Overview of the Genetic Structure within the Italian Population from Genome-Wide Data. Plos One 7: e43759.

16. Development Core Team R R (2009) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,Vienna, Austria. http://www.R-project.org.

17. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, et al. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28: 3326–3328.

18. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81: 559–575.

20. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461: 489–494.

21. Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583–589.

22. Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. Genome Res 23: 1514–1521.

23.Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655–1664.

24.McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, et al.(2008) Runs of homozygosity in European populations. Am J Hum Genet 83:359–372.

25.Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78: 629–644.

26.Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837.

27.Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4: e72.

28.Gautier M, Vitalis R (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28: 1176–1177.

29.Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 29: 308–311.

30. Vona G (1997) The peopling of Sardinia (Italy): History and effects. Int J Anthropol 12: 71–87.

31. Vona G, Floris-Masala R, Mameli GE, Succa V (1992) Red cell and serum protein polymorphisms in Sardinia. International Journal of Anthropology 7:25–33.

32. Piazza A, Cappello N, Olivetti E, Rendine S (1988) A genetic history of Italy. Annals of Human Genetics 52: 203–213.

33. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes.

34. Contu L, Carcassi C, Orru` S, Mulargia M, Arras M, et al. (1998) HLA-B35 frequency variations correlate with malaria infection in Sardinia. Tissue Antigens 52: 452–461.

35. Grimaldi MC, Crouau-Roy B, Amoros JP, Cambon-Thomsen A, Carcassi C, et al. (2001) West Mediterranean islands (Corsica, Balearic islands, Sardinia) and the Basque population: Contribution of HLA class I molecular markers to their evolutionary history. Tissue Antigens 58: 281–292.

36. 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.Human Molecular Genetics 9: 2959–2965.

37. Calo ` CM, Varesi L, Memmi M, Moral P, Vona G (2003) A pentanucleotide repeat polymorphism (TTTTA) in the apolipoprotein (a) gene - Its distribution and its association with the risk of cardiovascular disease. Collegium Antropologicum 27: 105–115.

38. Moral P, Marogna G, Salis M, Succa V, Vona G (1994) Genetic data on Alghero population (Sardinia): Contrast between biological and cultural evidence. American Journal of Physical Anthropology 93: 441–453.

39. Rendine S, Calafell F, Cappello N, Gagliardini R, Caramia G, et al. (1997) Genetic history of cystic fibrosis mutations in Italy. I. Regional distribution. Ann Hum Genet 61: 411–424.

40. Falchi A, Giovannoni L, Calo CM, Piras IS, Moral P, et al. (2006) Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. Journal of Human Genetics 51: 9–14.

41. Barbujani G, Bertorelle G, Capitani G, Scozzari R (1995) Geographical structuring in the mtDNA of Italians. Proceedings of the National Academy of Sciences of the United States of America 92: 9171–9175.

42. Fraumene C, Belle EMS, Castrı` L, Sanna S, Mancosu G, et al. (2006) High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. Molecular Biology and Evolution 23: 2101–2111.

43. Malaspina P, Cruciani F, Santolamazza P, Torroni A, Pangrazio A, et al. (2000) Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. Annals of Human Genetics 64: 395–412.

44. Morelli L, Grosso MG, Vona G, Varesi L, Torroni A, et al. (2000) Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. Human Biology 72: 585–595.

45. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. American Journal of Human Genetics 67: 1251–1276.

46. Capelli C, Redhead N, Romano V, Calı` F, Lefranc G, et al. (2006) Population structure in the Mediterranean basin: A Y chromosome perspective. Annals of Human Genetics 70: 207–225.

47. 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. American Journal of Physical Anthro-pology 121: 270–279.

48. 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. Human Immunology 62: 871–884.

49. 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.

50. 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. European Journal of Human Genetics 11: 802–807.

51. Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, et al. (2013) Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science 341: 565–569.

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

53. Naitza S, Porcu E, Steri M, Taub DD, Mulas A, et al. (2012) A Genome-Wide Association Scan on the Levels of Markers of Inflammation in Sardinians Reveals Associations That Underpin Its Complex Regulation. PLoS Genet 8:e1002480.

54 Pilia G, Chen WM, Scuteri A, Orru´ M, Albai G, et al. (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genetics 2:1207–1223.

55. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, et al. (2007) Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. Plos Genetics 3: 1200–1210.

56. Terracciano A, Sanna S, Uda M, Deiana B, Usala G, et al. (2010) Genome-wide association scan for five major dimensions of personality. Molecular Psychiatry 15: 647–656.

57. Sutin AR, Milaneschi Y, Cannas A, Ferrucci L, Uda M, et al. (2011) Impulsivity-related traits are associated with higher white blood cell counts. J Behav Med.

58. 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. Human Genetics 111: 9–15.

59. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. PLoS One 5: e13996.

60. Wagner ML (1941) Historische Lautlehre des Sardischen.

61. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19: 826–837.

62. Gompert Z, Buerkle CA (2011) A hierarchical Bayesian model for next-generation population genomics. Genetics 187: 903–917.

63. Gilad Y, Bustamante CD, Lancet D, Paabo S (2003) Natural selection on the olfactory receptor gene family in humans and chimpanzees. Am J Hum Genet73: 489–501.

64. Lopez Herraez D, Bauchet M, Tang K, Theunert C, Pugach I, et al. (2009) Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. PLoS One 4: e7888.