

**DOTTORATO DI RICERCA IN ETOLOGIA  
ECOLOGIA, ANTROPOLOGIA E BIOSISTEMATICA  
(XXVI CICLO)**

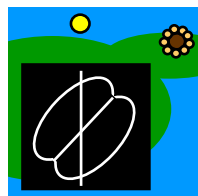
**STUDI DIACRONICI SUL DNA  
MITOCONDRIALE DELLE POPOLAZIONI  
ITALIANE: GLI UMBRI PLESTINI**

**Tesi di  
Alessandra Modi**



**Coordinatore Prof. Alberto Ugolini  
Tutor Prof. David Caramelli**

**(2013)**



*Ai miei genitori  
A Giulio*

# INDICE

<b>1. <u>INTRODUZIONE</u></b>	pag. 5
1.1 SULLE TRACCE DELLA PREISTORIA UMANA ATTRAVERSO LA GENETICA	pag. 7
1.1.1 I principali marcatori molecolari	pag. 8
1.2 IL DNA MITOCONDRIALE COME STRUMENTO PER LO STUDIO DELLA POPOLAZIONI UMANE ANTICHE E MODERNE	pag. 8
1.2.1 Polimorfismi del mtDNA e aplogruppi mitocondriali	pag. 9
1.2.2 Relazioni filogenetiche tra aptipi mitocondriali	pag. 11
1.2.3 Studi filogeografici	
1.2.4 L'analisi del fondatore	
1.3 ORIGINE DELLA STRUTTURA GENETICA DELLA POPOLAZIONE EUROPEA	pag. 13
1.3.1 Colonizzazione dell'Europa da parte dell'uomo anatomicamente moderno e LGM	pag. 14
1.3.2 L'avanzata della cultura neolitica	pag. 15
1.3.3 Contributo paleolitico Vs Contributo neolitico	pag. 16
1.3.4 Agricoltori neolitici	pag. 21
1.3.5 Espansioni post-neolitiche	pag. 23
1.4 IL CASO DELL'ITALIA	pag. 24
1.4.1 Neolitizzazione della penisola italiana	pag. 25
1.4.2 Le origini del patrimonio genetico italiano	pag. 26
1.5 GLI ABITANTI DELL'ITALIA ANTICA	pag. 27
1.5.1 Gli Umbri Plestini	pag. 29
1.5.2 I Messapi	pag. 30
1.5.3 I Sardi	pag. 31
1.6 IL DNA ANTICO: 30 ANNI DI STORIA	pag. 33
1.6.1 Principali difficoltà nello studio dell'aDNA	pag. 34
1.6.2 Tecniche di analisi del DNA antico	pag. 38

<b>2. <u>SCOPO DEL LAVORO</u></b>	pag. 45
<b>3. <u>MATERIALI E METODI</u></b>	pag. 46
3.1 CAMPIONAMENTO DEI REPERTI ANTICHI	pag. 46
3.1.1 Campioni Umbri	pag. 46
3.1.2 Campioni Pugliesi	pag. 46
3.1.3 Campioni Sardi	pag. 50
3.2 METODI	pag. 52
3.2.1 Preparazione del campione	pag. 53
3.2.2 Estrazione del DNA	pag. 54
3.2.3 Amplificazione dell'HVR1 mediante PCR	pag. 54
3.2.4 Preparazione delle librerie	pag. 56
3.2.5 Sequenziamento delle librerie	pag. 58
3.2.6 Mappaggio delle reads e ricostruzione della sequenza consenso	pag. 60
3.3 ANALISI DEI DATI	pag. 62
3.3.1 Analisi a livello intra-popolazionistico	pag. 62
3.3.2 Analisi a livello inter-popolazionistico	pag. 65
<b>4. <u>RISULTATI</u></b>	pag. 68
4.1 SARDEGNA	pag. 69
4.2 PUGLIA	pag. 72
4.3 UMBRIA	pag. 73
4.3.1 Analisi intra-popolazionistiche	
4.3.2 Analisi inter-popolazionistiche	pag. 77
<b>5. <u>DISCUSSIONE E CONCLUSIONI</u></b>	pag. 88
5.1 CONSIDERAZIONI FILOGENETICHE	pag. 88

5.2 CONSIDERAZIONI METODOLOGICHE

pag. 90

**6. BIBLIOGRAFIA**

pag. 91

**APPENDICE 1**

pag. 107

**APPENDICE 2**

pag. 121

**APPENDICE 3**

pag. 129

## RIASSUNTO

Per diverso tempo è stato assunto che il pool genetico dell'Europa attuale fosse principalmente il risultato delle migrazioni avvenuto durante il Neolitico. I recenti dati sull'aDNA in combinazione con le analisi descrittive e i modelli di simulazione computerizzati, hanno fatto rivalutare questa visione semplicistica. Emerge così che, in Europa, sono almeno tre gli eventi che più di tutti hanno agito sui patterns di diversità genetica: (i) la prima colonizzazione del continente da parte dell'uomo anatomicamente moderno attorno ai 45 ka; (ii) l'ultimo picco glaciale tra i 27 e i 16 ka e (iii) l'arrivo della cultura neolitica nel sudest Europa e la successiva espansione nel resto del continente tra i 9 e i 5 ka.

L'Italia, grazie alle sue particolari caratteristiche geografiche e climatiche, ha occupato sin dalla preistoria una posizione centrale nel flusso di scambi tra antiche popolazioni. Le numerose espansioni, dispersioni e immigrazioni hanno lasciato una traccia indelebile nel complesso mosaico sia culturale che genetico dell'Italia attuale: un metodo per studiare gli effetti di questi cambiamenti sulle dinamiche popolazionistiche è quello di analizzare la variazione di sequenza del DNA estratto dai reperti umani antichi. Con l'obiettivo di approfondire le conoscenze circa l'origine dell'attuale struttura genetica della popolazione italiana, abbiamo sequenziato la regione HVR1 del DNA mitocondriale di 103 campioni suddivisi tra le regioni Umbria (Umbri Plestini), Puglia (Messapi) e Sardegna (pre-nuragici), distribuiti in un arco temporale che va dal Neolitico Medio all'Età del Ferro. Come strategia sperimentale abbiamo utilizzato un approccio combinato metodologia classica/Next Generation Sequencing (454 *Genome Sequencer* – Roche), mentre per il mappaggio delle reads e la ricostruzione della sequenza consenso, abbiamo prodotto uno script specifico per l'analisi degli ampliconi.

Le analisi filogenetiche sono state condotte solo sugli Umbri Plestini, poiché Puglia e Sardegna hanno avuto una resa troppo bassa, con rispettivamente 9 e 6 sequenze complete. I polimorfismi di sequenza ci hanno consentito di associare le 30 sequenze plestine alle rispettive linee mitocondriali, individuando 26 aplotipi suddivisi in 12 hg dei quali l'H è quello maggiormente rappresentato, con il 37% della frequenza.

Per capire come si collocano gli Umbri Plestini nel contesto italiano ed europeo, sono state costruite le matrici di distanze genetiche con il modello Kimura 2-parametri, rappresentate poi graficamente con il metodo Multidimensional Scaling. Sia il confronto dei Plestini con le popolazioni italiane antiche sia quello con le popolazioni italiane ed eurasiatiche moderne, mettono in evidenza la peculiarità genetica dei Plestini che si discostano da tutti i gruppi antichi e moderni. Pare inoltre che essi non abbiano lasciato traccia nel pool genico attuale, neppure dei moderni abitanti della medesima regione.

## 1. INTRODUZIONE

La capacità di manipolazione e sfruttamento dell'ambiente sono uno dei tratti principali che caratterizzano la specie umana sin dalle sue origini: tracce di queste attività si trovano ovunque attorno a noi. Alcune deliberate come i megaliti di Stonehenge o le piramidi d'Egitto, altre decisamente più celate ma d'certo non meno importanti. Numerosissimi sono i siti archeologici che restituiscono resti fossili sia umani che animali, nonché tracce di insediamenti, avanzi di cibo e utensili. Tutte queste evidenze archeologiche ci forniscono importanti informazioni sulla dinamiche popolazionistiche che hanno interessato il nostro continente e testimoniano la vita quotidiana delle società preistoriche. Inoltre, grazie all'analisi stratigrafica e alle datazioni dirette, è possibile collocare in un contesto cronologico tali resti archeologici.

E' proprio grazie alle testimonianze archeologiche e alle datazioni al radiocarbonio, che è stato possibile identificare la prima colonizzazione dell'Europa da parte dell'uomo anatomicamente moderno (AMH, Anatomically Modern Human): fino a poco tempo fa, la prima comparsa di AMH era stata fissata a circa 42-43 ka, basandosi unicamente sull'associazione con i manufatti aurignaziani (Mellars, 2006; Higham et al., 2011). Recenti datazioni al radiocarbonio sui reperti fossili hanno confermato questi dati collocando la comparsa di AMH in Europa tra i 44,2-41.5 calibrati (cal.) ka BP nella Kent's Cavern nel sud dell'Inghilterra (Higham et al., 2011a) e tra i 45-43 cal. ka BP presso Grotta del Cavallo in Italia (Benazzi et al., 2011), condividendo gran parte degli habitat con i Neandertaliani – anche se questi ultimi non sono sopravvissuti oltre i 39 cal. ka BP in molte regioni dell'Europa e del Caucaso (Higham, 2001b; Pinhasi et al., 2011).

Il record archeologico testimonia come il più grande cambiamento nelle società umane inizi attorno ai 10.000 anni fa nell'area della Mezzaluna Fertile (Childe, 1925): è qui che troviamo i primi segni della domesticazione di piante e animali (Harris, 1996; Clutton-Brock, 1999). Questa transizione da società di cacciatori-raccoglitori a società di agricoltori viene comunemente indicata come "Neolitizzazione". Con essa sono avvenuti cambiamenti importanti, quali la comparsa di insediamenti stabili e successivamente villaggi, lo sviluppo di strutture sociali più organizzate, un rapido progresso tecnologico, come pure l'insorgere di guerre e nuove malattie (Diamond, 2002). Dal vicino Oriente questo nuovo stile di vita si è poi rapidamente diffuso in Europa sino a 5.000 anni fa (Ammerman & Cavalli-Sforza, 1984; Pinhasi et al., 2005; Fort et al., 2012).

Anche se, come abbiamo appena visto, la documentazione archeologica ci è di grande aiuto nel capire la storia del passato, tuttavia non è in grado di rispondere ad alcune importanti domande sulla nostra origine: da dove provenivano i primi colonizzatori dell'Europa? Come hanno interagito con i Neandertaliani? In che modo lo stile di vita neolitico è arrivato sino a noi? In questi casi un ulteriore strumento d'indagine per studiare le dinamiche popolazionistiche del passato è fornito dalle analisi genetico-molecolari. Studi condotti sulle popolazioni attuali hanno messo in evidenza come l'intera umanità discenda da un progenitore africano (o per meglio dire, da una progenitrice africana, dal momento che questi studi sono stati condotti sul DNA mitocondriale, mtDNA) comparso attorno ai 200.000 anni fa (Cann et al., 1987; Ingman et al., 2000), così come analisi condotte sulle moderne specie domestiche hanno dato un'idea circa l'antichità e il numero di eventi di domesticazione che si sono realizzati nel corso del tempo (Vilà et al., 1997; Luikart et al., 2001; Savolainen et al., 2002). Eppure, quando si utilizzano dati ottenuti da popolazioni moderne per ricostruire eventi storici, c'è il rischio di incorrere in interpretazioni errate dal momento che le dinamiche demografiche recenti possono aver oscurato gli eventi più antichi. Questo problema può essere superato studiando il DNA recuperato dai reperti fossili antichi. Ne sono esempi lampanti i recenti studi condotti su campioni neandertaliani e AMH: il confronto di due genomi preistorici, uno ottenuto da tre individui neandertaliani (Green et al., 2010) e uno da un ominide siberiano non ancora caratterizzato da un punto di vista morfologico (Reich et al., 2010; Reich et al., 2011), stanno rivalutando la rigorosa origine africana dell'uomo anatomicamente moderno nonché le tanto dibattute interazioni *sapiens*-Neanderthal, dal momento che da entrambi gli studi emerge che AMH possa essersi ibridato con le specie arcaiche europee.

Lo studio del DNA antico (aDNA) ha inoltre fornito importanti informazioni all'altro storico dibattito: quello sulla diffusione della cultura neolitica. Su questo argomento si sono da sempre fronteggiate tre teorie; la diffusione del Neolitico è stata un processo (i) prevalentemente demico (Childe, 1925); (ii) prevalentemente culturale (Edmonson, 1961) oppure (iii) prevalentemente demico in alcune regioni e prevalentemente culturale in altre (Zeder, 2008). Lo studio degli aplogruppi e le simulazioni di coalescenza tra sequenze mitocondriali neolitiche e moderne hanno dimostrato che in molte regioni dell'Europa la transizione neolitica ha seguito un processo prevalentemente demico (Gamba et al., 2012; Skoglund et al., 2012; Sanchez-Quinto et al., 2012; Brotherton et al., 2013), dal momento che non c'è continuità genetica tra le popolazioni mesolitiche e quelle neolitiche, ma che anzi, queste ultime mostrano forte affinità con le popolazioni del vicino Oriente.

Da trenta anni a questa parte, quindi, la capacità di recuperare e studiare il DNA da reperti fossili ha reso possibile uno studio più accurato delle popolazioni antiche, dei loro movimenti e delle interazioni tra loro e l'ambiente



circostante, mettendo in luce scenari assolutamente inimmaginabili con il solo contributo dell'archeologia.

## **1.1 SULLE TRACCE DELLA PREISTORIA UMANA ATTRAVERSO LA GENETICA**

L'attuale struttura genetica dell'uomo è il risultato della complessa combinazione tra processi demografici, selezione naturale ed eventi stocastici, che hanno plasmato il genoma lasciando su di esso tracce indelebili delle loro azioni.

In una popolazione la diversità molecolare è dovuta alle mutazioni e al flusso genico seguiti dalla deriva e dalla selezione e si manifesta come sostituzioni nucleotidiche, inserzioni e delezioni. L'ammontare e il pattern di variazione sono influenzati dai processi evolutivi combinati alla storia demografica della popolazione stessa. Grazie alle diverse modalità di trasmissione ereditaria, il DNA mitocondriale, il cromosoma Y e i marcatori del DNA nucleare ci consentono di ripercorrere la storia umana attraverso una linea, rispettivamente, materna, paterna o di entrambi i sessi.

Per visualizzare la variazione genetica e la sua distribuzione nello spazio, si utilizzano solitamente networks e alberi filogenetici, mentre le analisi di coalescenza ci consentono di ripercorrere a ritroso la storia genetica di un campione sino al più recente antenato comune (MRCA). Stimando i tassi di mutazione e assumendo specifici modelli di insorgenza delle mutazioni, è possibile calcolare il tempo di divergenze delle popolazioni umane: il limite di questa procedura però, è che ci sono ancora diverse incertezze sulle stime dell'orologio molecolare (Ho & Larson, 2006; Fu et al., 2013). Tuttavia, sappiamo con sicurezza che specifiche caratteristiche della diversità genetica sono il risultato di differenti eventi demografici. Esempio: si assume che il centro di origine di una popolazione corrisponde a quella regione dove si riscontra la più alta variabilità genetica tra gli individui; l'antichità di quella particolare popolazione infatti, ha permesso l'accumulo di un maggior numero di mutazioni. Al contrario, quando le popolazioni passano attraverso colli di bottiglia – come per esempio le migrazioni di piccoli gruppi di persone oppure la selezione di certi gruppi animali per scopi domestici – si ha una netta riduzione della diversità genetica (Nei et al., 1975). Se questi gruppi (derivanti da un'unica popolazione d'origine) si espandono rapidamente, il network si presenterà con una tipica struttura a stella (Slatkin & Hudson, 1991) mentre l'eccesso di alleli rari aumenterà il grado di eterozigosi e ridurrà il linkage disequilibrium (Watterson, 1986). Allo stesso modo, se una popolazione di piccole dimensioni ha un basso grado di admixture, la deriva genica causa perdita allelica e decremento della diversità genetica (Kimura, 1955).

E' quindi possibile stimare quantitativamente e qualitativamente gli eventi demografici che hanno plasmato la costituzione genetica delle attuali popolazioni umane grazie all'analisi del DNA combinata agli studi di simulazione mediante modelli statistici in campioni antichi e moderni.

### **1.1.1 I principali marcatori molecolari**

A livello genomico, la variabilità di un gruppo di individui può essere studiata sfruttando diversi marcatori molecolari, ma non tutti possono essere impiegati su campioni antichi data la natura del DNA che si estrae dai fossili. Ad esempio, seppure siano stati utilizzati per indagare la storia e l'evoluzione delle popolazioni, delle specie ecc. (Ellegren, 2004), i microsatelliti del DNA nucleare non sono marcatori affidabili in paleogenetica dal momento che richiedono sequenze troppo lunghe per un genoma antico e degradato. Per questo tipo di campioni sono decisamente più idonei i polimorfismi a singolo nucleotide (SNPs) poiché si possono studiare attraverso corte sequenze di DNA, siano esse nucleari, mitocondriali o Y-cromosomiche.

Sia l'mtDNA che il DNA Y-cromosomico sono sistemi aploidi in cui una particolare combinazione di caratteristiche genetiche è definita come aplotipo (ht) e dove un gruppo monofiletico di aplotipi sono definiti come aplogruppo (hg). Il DNA mitocondriale è particolarmente adatto per gli studi sull'aDNA e per indagare i modelli filogeografici recenti grazie alle sue peculiari caratteristiche: poliplasmia, alto tasso di mutazione e modalità non ricombinante di eredità. Soprattutto la regione di controllo, altamente polimorfica, con le sue regioni ipervariabili I e II (Vigilant et al., 1991; Lutz et al., 1998) è da sempre la più utilizzata per l'identificazione degli aplogruppi e negli studi filogenetici.

## **1.2 IL DNA MITOCONDRIALE COME STRUMENTO PER LO STUDIO DELLE POPOLAZIONI UMANE ANTICHE E MODERNE**

A partire da gli anni '80, lo studio della variabilità del mtDNA è diventato uno strumento importante per ricostruire la storia delle popolazioni umane. I primi a utilizzare l'mtDNA come marcatore molecolare per la ricostruzione del passato recente furono Westley Brown e Douglas Wallace alla fine degli anni '70. Storicamente l'mtDNA veniva caratterizzato mediante polimorfismi di restrizione enzimatica (RLFPs, Restriction Fragments Length Polymorphisms), per mezzo dei quali le diverse varianti mitocondriali sono

identificate in base alla presenza/assenza di precise posizioni di restrizione. Nei primi studi, a bassa risoluzione, campioni popolazionistici discretamente ampi venivano analizzati utilizzando uno o pochi enzimi di restrizione (Brown, 1980; Denaro et al., 1981), mentre la tendenza negli studi successivi fu quella di utilizzare numerosi enzimi di restrizione su un elevato numero di campioni (Jonhson et al., 1983; Cann et al., 1987; Santachiara-Benerecetti et al., 1988; Scozzari et al., 1988; Torroni et al., 1994, 1996). Tuttavia, questo approccio catturò l'attenzione pubblica solamente nel 1987, quando Rebecca Cann, Mark Stoneking e Allan Wilson utilizzando 12 enzimi, riuscirono ad ottenere un'accurata filogenesi dell'mtDNA. La loro analisi di 147 mtDNA provenienti da differenti continenti identificò ciò che fu impropriamente definita come "l'Eva mitocondriale", e portò all'ipotesi che tutti gli mtDNA moderni discendano da quello di una donna che visse in Africa circa 200.000 anni fa (Cann et al., 1987).

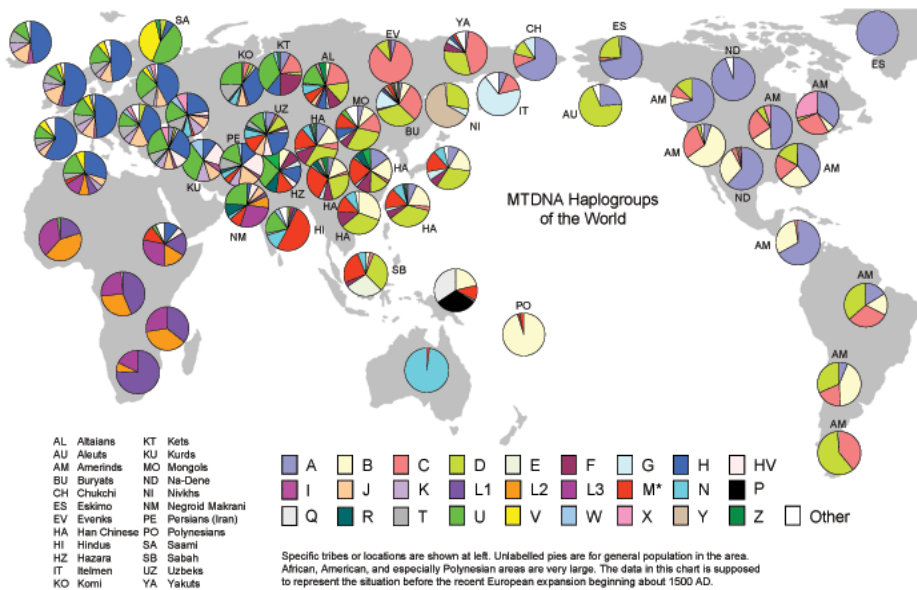
Agli inizi degli anni '90 l'analisi dell'mtDNA, in parallelo con gli studi che si occupavano dell'origine di *Homo sapiens*, fu applicata su campioni molto ampi di soggetti provenienti da singoli continenti, con l'obiettivo di determinare l'origine delle popolazioni umane presenti in ogni principale area geografica. Questo condusse ad un quadro molto più definito della filogenesi mondiale dell'mtDNA. In seguito lo sviluppo e la diffusione della PCR (Polymerase Chain Reaction) e l'avanzamento delle tecniche di sequenziamento hanno permesso un nuovo approccio molecolare, ovvero lo studio di sequenza di porzioni di mtDNA: la risoluzione di questa nuova tecnica è decisamente superiore rispetto alle analisi con enzimi di restrizione e ha consentito un'indagine più dettagliata dei singoli tipi mitocondriali. Inizialmente le analisi erano incentrate sulla sola D-loop, mentre in seguito, con il continuo miglioramento delle metodologie d'indagine, sono state estese ad altre regioni non codificanti sino ad arrivare al sequenziamento di interi genomi mitocondriali, sia moderni che antichi di specie diverse, mediante Next Generation Sequencing (NGS) (Gilbert et al., 2008; Briggs et al., 2009; Edwards et al., 2010; Krause et al., 2010; Lari et al., 2011; Knapp et al., 2012; Sanchez-Quinto et al., 2012; Brotherton et al., 2013; Dabney et al., 2013; Fu et al., 2013b).

La combinazione dei risultati degli enzimi di restrizione su regioni codificanti e non, con i dati di sequenza ha portato ad una classificazione sempre più articolata dei diversi aplogruppi mondiali che si riflette anche su una nomenclatura piuttosto complessa.

### **1.2.1 Polimorfismi del mtDNA e aplogruppi mitocondriali**

Il concetto di aplogruppo è stato introdotto per la prima volta da Torroni, Wallace e colleghi all'inizio degli anni novanta (Torroni et al., 1992).

Si definisce aplogruppo o cluster mitocondriale un insieme di aplotipi che condividono specifiche sostituzioni nucleotidiche diagnostiche e che hanno un'origine comune. Ciò implica che i polimorfismi che identificano ciascun aplogruppo si sono verificati nelle linee ancestrali di tutti gli aplotipi che lo costituiscono. La filogenesi mondiale dell'mtDNA mostra come i principali aplogruppi siano geograficamente circoscritti, così che si possono distinguere dei cluster tipicamente africani, europei e asiatici (Fig. 1.1): questa caratteristica è risultata di estrema utilità per l'inferenza di modelli migratori intra- ed extra-continentali.



**Fig 1.1:** distribuzione geografica degli aplogruppi mitocondriali mondiali.

I diversi aplogruppi sono indicati con una nomenclatura piuttosto complessa: inizialmente Wallace e Torroni (Torroni et al., 1992) proposero di identificare i principali cluster con una lettera maiuscola (esempio: A, B, C, D, H, I, L, U, V). Ogni cluster può essere a sua volta diviso in un sotto-aplogruppo che si designa con la lettera corrispondente seguita da un numero (esempio: L2). La suddivisione a livello successivo si realizza alternando lettere minuscole e numeri (esempio: U5a1b). Le linee che presentano il motivo tipico di un aplogruppo ma diverso motivo per il sotto-aplogruppo si indicano con un asterisco e solitamente prendono il nome di paragrappi anziché aplogruppi

(esempio: J\*). Quando due o più aplogruppi condividono una serie di motivi che suggeriscono un'origine comune, si tende a raggrupparli unendo i nomi dei differenti cluster (esempio: HV, JT).

### **1.2.2 Relazioni filogenetiche tra aplotipi mitocondriali**

Le analisi sulla regione HVR1 hanno evidenziato che tale segmento è caratterizzato da un tasso di mutazione diverso lungo la sequenza: mentre alcune posizioni sono estremamente variabili tra i diversi gruppi umani, altre sono invece relativamente costanti (Piercy et al., 1993; Wakeley, 1993). Una medesima sostituzione nucleotidica può essere presente in popolazioni diverse per due ragioni: perché viene ereditata dai discendenti a partire da un antenato ancestrale, oppure perché si è realizzata indipendentemente in linee differenti per evoluzione convergente (omoplasia). Qualsiasi analisi filogenetica che dà lo stesso peso a tutte le sostituzioni, tenderà a raggruppare le sequenze che condividono le medesime mutazioni, creando gruppi polifiletici: ogni interpretazione evolutiva che deriva da queste analisi sarà quindi errata. Analogamente, le sostituzioni multiple di una data posizione e le retromutazioni portano ad una sottostima del numero di sostituzioni che hanno effettivamente avuto luogo nel corso dell'evoluzione: in questo caso risulterà difficile, se non impossibile, determinare lo stato ancestrale dell'aplotipo e calcolare la distanza genetica tra la sequenze.

A causa di queste problematiche, man mano che aumentavano i dati sulla variabilità dell'mtDNA delle popolazioni umane, i “metodi classici” di ricostruzione filogenetica diventavano sempre meno efficaci ed urgeva lo sviluppo di nuovi strumenti di analisi per lo studio di questo particolare materiale genetico. L'approccio più efficace è stato quello di Bandelt et al. 1995: questo metodo di ricostruzione filogenetica – chiamato “Median Networks” – permette di riassumere in un'unica rete gli alberi più parsimoniosi generati da un gruppo di dati, evidenziando anche la frequenza di ciascuno degli aplotipi osservati. Quando il numero di aplotipi differenti è molto elevato, si generano reticoli iperdimensionali molto complessi che rendono l'albero di difficile interpretazione: in questo caso si usa eliminare gli aplotipi meno frequenti partendo dal presupposto che sono quelli che hanno un'origine più recente (Richards et al., 1998). Un metodo alternativo per la costruzione di networks con poche reticolazioni è il “Median Joining” (Bandelt et al., 1999): in questo caso vengono inseriti dei nodi nella rete di sequenze osservate che identificano l'aplotipo ancestrale più probabile.

La topologia dei networks fornisce informazioni sulla storia demografica di una popolazione. Per esempio, un network a forma di stella – con un aplotipo centrale dal quale si dipartono numerosi aplotipi separati da quello più antico da un singolo evento mutazionale – indica un'espansione popolazionistica. In

questo caso, una determinata variante mitocondriale è soggetta a condizioni riproduttive favorevoli per un lungo periodo di tempo e aumenta di frequenza. Dopo diverse generazioni alcuni dei discendenti portatori di questo aplotipo sviluppano nuove mutazioni e, in alcuni casi, la variante ancestrale si può estinguere in seguito a eventi di deriva genica. Se qualcuno dei discendenti va in contro ad un processo analogo, nella filogenesi compaiono i sottogruppi.

### **1.2.3 Studi filogeografici**

Integrando i dati di variabilità genetica con la distribuzione geografica, è possibile valutare l'impatto dei processi storici sulla diversità genetica attuale (Avice et al. 1987; Templeton et al., 1995). Le inferenze filogeografiche richiedono delle stime molto accurate sia della variazione spaziale che di quella temporale dei marcatori molecolari.

Esistono diversi metodi per rappresentare la variabilità geografica delle frequenze aplotipiche. Templeton e collaboratori (Templeton et al., 1995; Posada et al., 2000) hanno sviluppato il metodo della "Nested Cladistic Phylogeographical Analysis" (NCPA), che valuta i dati filogeografici stimando la significatività statistica dell'associazione geografica degli aplotipi. In questo modo è possibile studiare eventi quali la dispersione, l'espansione o altri eventi storici delle popolazioni. Ma il metodo più semplice e tradizionalmente più utilizzato per gli aplogruppi mitocondriali è quello di creare delle "mappe di frequenza", riportando la frequenza osservata sopra una mappa o unendo i punti con ugual frequenza genica mediante una linea denominata "linea isogenica". Tuttavia questo approccio ha un forte limite dato dal fatto che su ogni mappa si possono riportare le informazioni di un solo allele. A tal proposito, Cavalli-Sforza e collaboratori hanno sviluppato un metodo che consente di rappresentare le informazioni di più alleli su un'unica mappa (Cavalli-Sforza et al., 1993): le informazioni di ciascun allele vengono condensate in un numero ridotto di variabili mediante il metodo delle Componenti Principali. Ogni componente principale si calcola come la funzione lineare ottimizzata di tutta la frequenza genica disponibile ed esprime la percentuale della variazione genetica osservata. Se si rappresenta su una mappa la frequenza genica utilizzata per la costruzione di ciascuna componente principale, si ottiene una "mappa sintetica" di ciascuna di esse. Sulla mappa, la presenza di un gradiente a forma di anello concentrico attorno ad un'area circoscritta si interpreta come una espansione demografica con centro d'origine in quella regione (Cavalli-Sforza et al. 1993).

### **1.2.4 L'analisi del fondatore**

L'analisi del fondatore consente di identificare e datare la colonizzazione di nuovi territori, riconoscendo specifiche sequenze fondatrici in potenziali popolazioni d'origine e datando i clusters derivati da esse, presenti nella zona d'interesse (Richards et al., 2000). Allo stesso modo è possibile stimare il tempo di migrazione osservando la diversità accumulata dagli aplotipi fondatori nella nuova popolazione.

Questo tipo di analisi si realizza utilizzando la variabilità osservata nelle popolazioni attuali con le seguenti considerazioni:

- le popolazioni attuali sono discendenti delle popolazioni del passato e si trovano localizzate nella medesima regione geografica;

- la diversità genetica delle popolazioni attuali è molto simile a quella delle popolazioni del passato .

L'analisi del fondatore è stata utilizzata per identificare l'origine delle popolazioni del Pacifico (Stoneking & Wilson, 1989; Stoneking et al., 1990; Sykes et al. 1995) e di quelle amerinde (Torroni et al., 1993; Forsters et al., 1996) e per determinare l'origine, il tempo e l'impatto delle differenti migrazioni che hanno plasmato la struttura genetica dell'Europa.

### **1.3 ORIGINE DELLA STRUTTURA GENETICA DELLE POPOLAZIONI EUROPEE**

Per diverso tempo è stato assunto che la struttura geografica della variabilità genetica entro l'Europa fosse principalmente il risultato del flusso genico avvenuto durante e subito dopo la neolitizzazione. I recenti dati sull'aDNA ottenuti grazie alle tecniche di sequenziamento ultramassivo in combinazione con le analisi descrittive e con i modelli di simulazione computerizzati, hanno fatto rivalutare questa visione semplicistica. Emerge così che, in Europa, sono almeno tre gli eventi che più di tutti hanno agito sui patterns di diversità genetica: (i) la prima colonizzazione del continente da parte dell'uomo anatomicamente moderno attorno ai 45 ka; (ii) l'ultimo picco glaciale tra i 27 e i 16 ka e (iii) l'arrivo della cultura neolitica nel sud-est Europa e la successiva espansione nel resto del continente tra i 9 e i 5 ka.

### 1.3.1 Colonizzazione dell'Europa da parte dell'uomo anatomicamente moderno e LGM

I dati archeologici, cronologici e molecolari sono concordi nel collocare la prima colonizzazione dell'Europa da parte dell'AMH in un periodo compreso tra i 45 e i 41 ka BP (Mellars, 2006; Higham et al., 2011; Higham et al., 2011a; Benazzi et al., 2011) a partire dal vicino Oriente secondo le rotte evidenziate nella figura 1.2

Dopo la scomparsa dei Neandertal ed in particolare durante l'ultimo massimo glaciale (LGM, Last Glacial Maximum) gran parte del nord Europa era ricoperto dai ghiacci, così che le popolazioni umane furono costrette a ridurre il loro areale di distribuzione, ritirandosi in rifugi a sud-est del continente europeo, lungo il Mediterraneo, i Balcani e il Levante e nelle pianure dell'est (Anikovich et al., 2007; Pavlov et al., 2001; Jochim, 2011; Morin, 2008). Cambiamenti nei pattern di distribuzione della cultura materiale sono interpretati come variazioni della distribuzione geografica delle popolazioni: durante il Gravettiano si ha una colonizzazione delle regioni oltre i 52°N di latitudine a cui fa seguito un parziale o totale abbandono di questi territori attorno ai 24 ka per poi essere ricolonizzati tra i 20 e 16 ka; al contrario si ha una certa continuità occupazionale nelle regioni più meridionali (Verpoorte, 2009). Tuttavia, resta difficile stabilire in quale misura le diverse culture materiali corrispondono effettivamente a popolazioni distinte e quanto la loro variazione nella distribuzione spazio-temporale è correlata a processi demografici. Anche se è probabile che qualche componente degli europei moderni derivi da una o più popolazioni rifugiate durante il LGM, il numero, la collocazione e la distribuzione di questi rifugi rimangono incerti (Underhill & Kivisild, 2007; Nielsen & Beaumont, 2009; Soares et al., 2010; Stewart & Stringer, 2012).

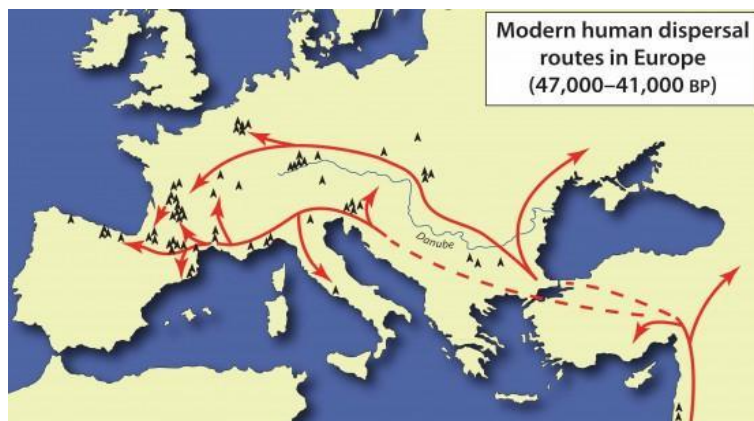
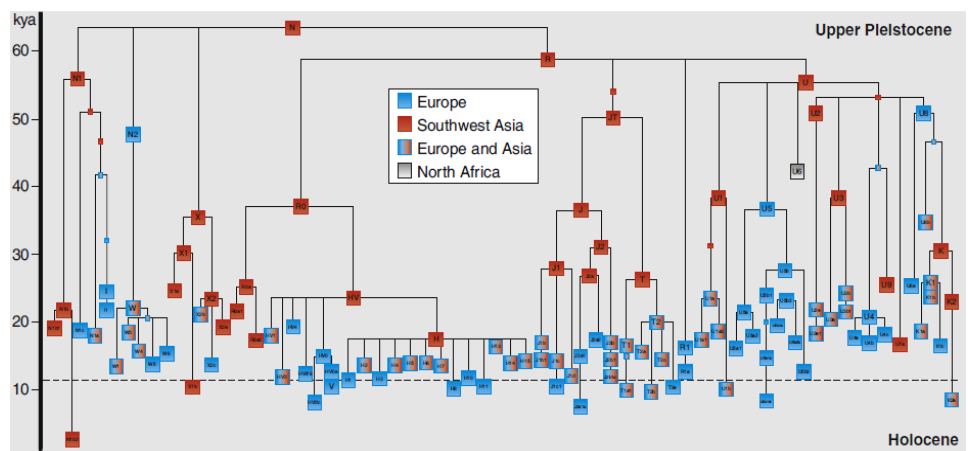


Fig 1.2: rotte migratorie seguite da AMH per la colonizzazione dell'Europa.



L' mtDNA delle moderne popolazioni europee mostra comunque evidenze di ri-espansione e ri-insediamento del centro e del nord Europa a seguito della fase più fredda, dopo i 15 ka: l'analisi del fondatore sugli aplogruppi oggi più diffusi in Europa – V, H1, H3, H5 e U5b1 (Fig. 1.3) – colloca la loro origine nel sud-est europeo. Da qui si sarebbero poi diffusi, con diverse possibili rotte migratorie, ad ovest, nel centro e nel nord Europa, nel corso del miglioramento climatico (Tambets et al., 2004; Gamble et al., 2005; Pereira et al., 2005; Torroni et al., 2006). L'aplogruppo H, il più diffuso nelle moderne popolazioni europee con una frequenza del 45%, pare sia apparso nel Vicino Oriente circa 18 ka. La comparsa in Europa è attualmente attestata a circa 15 ka, suggerendo un ingresso dopo il LGM. Ugualmente, l'aplogruppo V – un sister clade dell'H (Fig. 1.3) – ha un'ascendenza in Europa, ma più probabilmente si è diffuso da sud-est dopo il LGM (Torroni et al. 2001; Soares et al. 2010).



**Fig 1.3:** albero filogenetico dei principali aplogruppi mitocondriali presenti in Europa, sud-est asiatico e nord-Africa (immagine da Soares et al, 2010).

### 1.3.2 L'avanzata della cultura neolitica

Mentre in Europa si attraversa la fase “mesolitica”, le popolazioni oloceniche del Vicino Oriente rispondono ai cambiamenti ambientali in maniera piuttosto diversa: i Natufiani, nome con il quale vengono designate le popolazioni mesolitiche del Vicino Oriente (Sherratt, 1997), nel corso dell'ultima glaciazione diventano via via sempre più sedentarie, incrementando la raccolta del grano selvatico e, come qualcuno sostiene, a partire da 12 ka,

adottando la coltivazione dei cereali, processo quest'ultimo graduale e geograficamente disperso (Brown et al., 2008). Altri suggeriscono che l'agricoltura compare durante il Neolitico pre-ceramico A, successivamente ai Natufiani. Ad ogni modo, le piante domestiche sono attestate inequivocabilmente nel primo Neolitico pre-ceramico B a partire da 10.5 ka, a cui fa seguito la documentazione delle più antiche forme di domesticazione animale (Kuijt & Goring-Morris, 2002).

Su quali siano state le pressioni che hanno portato ad un così profondo cambiamento, il dibattito è ancora aperto: c'è disaccordo sul fatto che il Neolitico è stato una risposta economica necessaria per far fronte ad una crisi generatasi durante lo Younger Dryas (Bar-Yosef, 1998), o piuttosto la risposta a pressioni demografiche, o a cambiamenti religiosi e della cultura simbolica (Cauvin, 2000). Altra incognita, dov'è che si sono manifestati i primi segni di cambiamento: in Anatolia, nel Levante o nelle zone più a nord della Mezzaluna Fertile? La genetica non può fornire un contributo significativo al primo problema, mentre è in grado di far luce sull'origine geografica del Neolitico. Per esempio: le due principali linee dell'aplogruppo J del cromosoma Y nel Vicino Oriente sono strettamente correlate con lo sviluppo dell'agricoltura nelle aree più umide del nord (J2) e della pastorizia nelle aride regioni del sud (J1) (Chiaroni et al., 2008), anche se le più antiche evidenze archeologiche neolitiche si ritrovano nelle zone del sud (Kuijt & Goring-Morris, 2002). In ogni caso, J2 risulta essere il più importante marcatore Y-cromosomico per l'espansione dell'agricoltura nel sud-est Europa (Semino et al., 2000; di Giacomo et al., 2004).

I dati genetici ci forniscono informazioni importanti anche per capire come la cultura neolitica ha raggiunto l'Europa. Il modello dell' "ondata di diffuse" (o della "diffusione demica") sostiene che i clini in direzione sud-est/nord-ovest mostrati da diversi marcatori genetici riflettono in larga scala l'espansione degli agricoltori asiatici verso l'Europa. Dall'altro lato, numerosi archeologi ipotizzano uno sviluppo endogeno dell'agricoltura o comunque l'adozione della nuova cultura dal Vicino Oriente per mezzo di limitate e geograficamente ristrette, migrazioni umane (Ammerman & Cavalli-Sforza, 1973; Ammerman & Cavalli-Sforza, 1984; Chikhi et al. 1998; Berker, 2006).

### **1.3.3 Contributo paleolitico Vs Contributo neolitico**

Il primo studio mirato a ricostruire la storia europea basandosi sulla genetica risale a oltre trenta anni fa (Menozzi et al. 1978): utilizzando l'analisi delle componenti principale (PCA) vennero create delle mappe sintetiche riassumendo la variazione spaziale delle frequenze alleliche di 39 marcatori molecolari classici (come i gruppi sanguigni). La prima mappa mostrava un gradiente sud-est/nord-ovest che fu attribuito, come detto precedentemente, ad

una diffusione demica degli agricoltori verso l'Europa; i graduali cambiamenti nelle frequenze alleliche furono invece interpretati come l'introggressione di varianti paleolitiche locali nelle comunità agricole: durante il processo di diffusione, l'admixture tra cacciatori-raccoglitori locali e agricoltori, fu minimo. Di conseguenza, fu teorizzato che il pool genico degli attuali Europei consiste di una componente principale neolitica di origine anatolica (o comunque del Vicino Oriente) e di una piccolissima percentuale proveniente dalle popolazioni europee pre-neolitiche. Questo studio fu in seguito implementato utilizzando le frequenze alleliche di oltre 130 marcatori molecolari (Cavalli-Sforza et al., 1993) che produssero risultati simili.

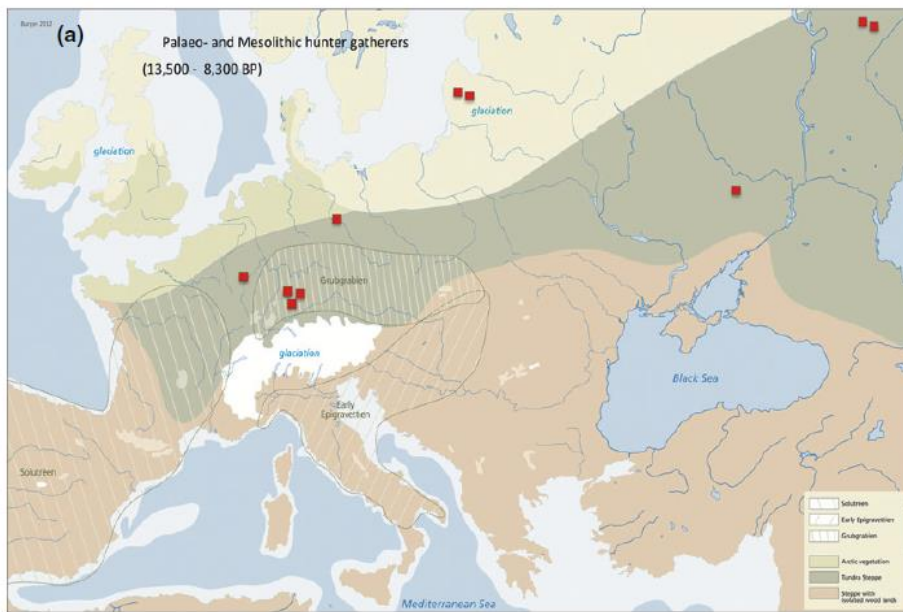
Successivamente, diversi altri lavori hanno utilizzato le analisi PCA per lo studio della struttura genetica delle popolazioni attuali (Fix, 1999; Sokal et al., 1999; Jobling et al., 2004; Novembre & Stephens, 2008; Francois et al., 2010). Esse infatti possono fornire sintesi utili di alcuni dati e affinare intuizioni quando si sviluppano ipotesi sulla storia delle popolazioni, ma comunque dovrebbero essere interpretate con cautela perché (i) non forniscono prove dirette delle ipotesi in via di sviluppo; (ii) non sempre sono rappresentativi dei processi demografici che descrivono (Novembre & Stephens, 2008; Francois et al., 2010); (iii) mostrano soltanto una porzione limitata di tutta la variazione genetica presente (per esempio: le prime due componenti principali della variazione genetica nel genoma europeo [Novembre et al., 2008] esprimono soltanto le 0.45% della variazione totale); e (iv) possono essere distorte da quantità e qualità del dato. Quindi, le PCA così come altri approcci di clustering, possono fornire interessanti sintesi dei dati e rivelare relazioni inaspettate tra le popolazioni, ma certamente non possono sostituire le analisi dirette sui processi demografici e sulle ipotesi evuzionistiche (Thomas et al., 2013).

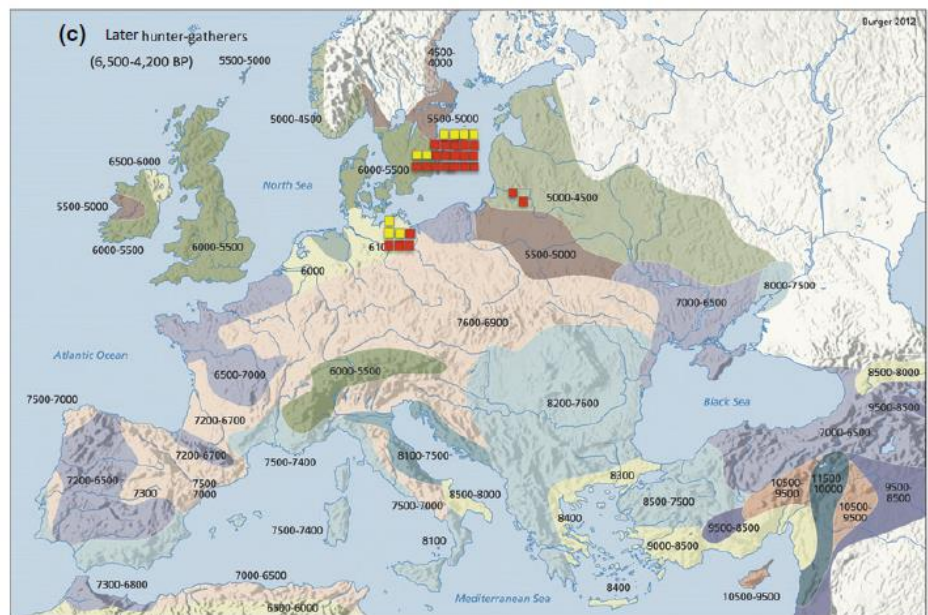
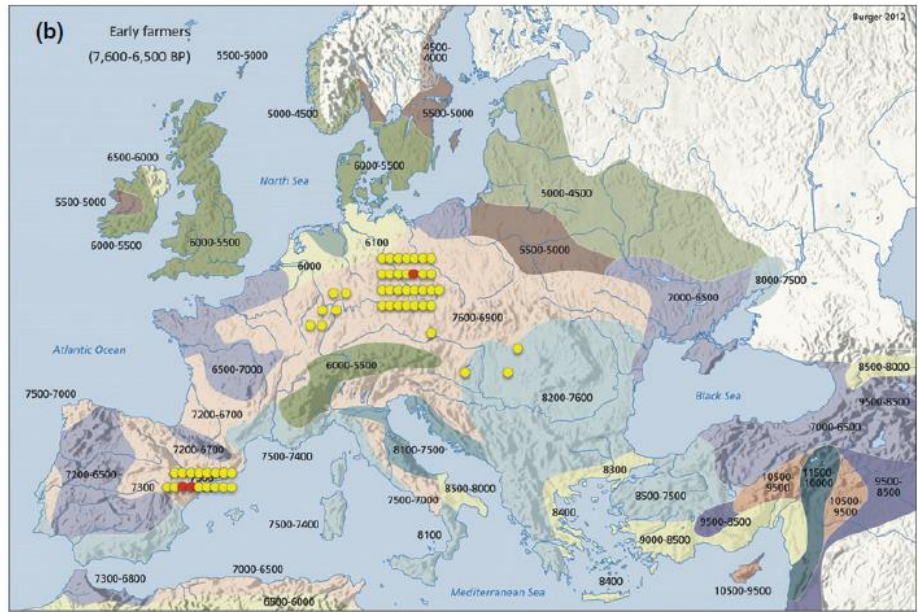
Un altro limite è che tutti questi studi sono stati condotti sulle popolazioni moderne: la diversità genetica attuale però, fornisce soltanto una testimonianza indiretta della storia evolutiva delle popolazioni antiche. Dai dati sul DNA moderno si evince che la similarità genetica riscontrata negli Europei attuali è caratterizzata da una forte correlazione con la geografia (Sokal, 1988; Novembre et al., 2008) ma, come detto precedentemente, la misura in cui questo pattern può essere attribuito ad una iniziale colonizzazione dell'Europa da parte di AMH, ad una ricolonizzazione dopo il ritiro dei ghiacci, alla transizione Neolitica, o ad un generale isolamento dovuto alle distanze territoriali, resta pressoché irrisolto. Lo studio dell'aDNA può aiutare a risolvere tali questioni fornendo dati diretti circa la diversità genetica delle popolazioni del passato.

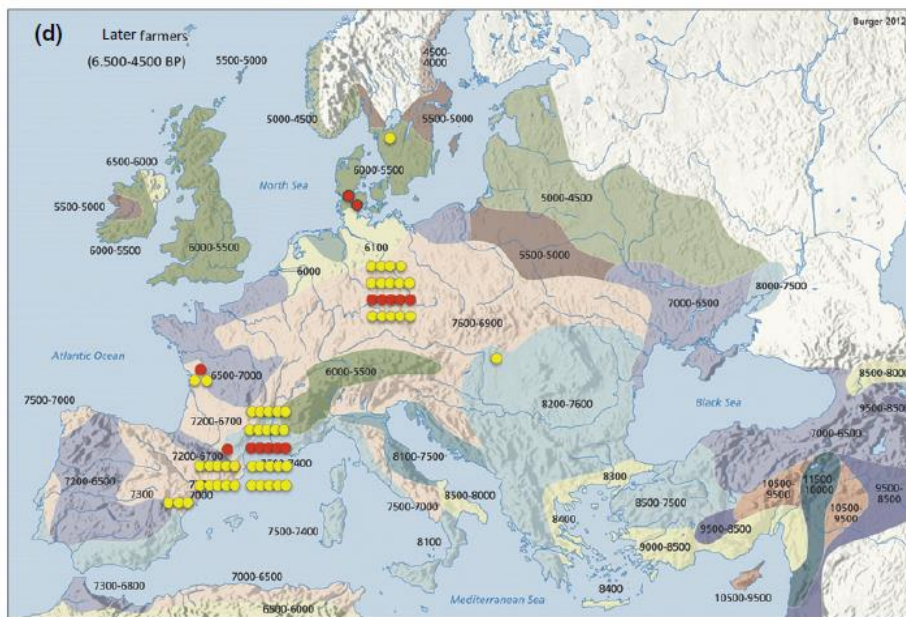
Le sequenze mitocondriali recuperate da tre individui del Paleolitico Superiore e da quattordici cacciatori-raccoglitori Mesolitici e Neolitici appartengono tutte all'haplog U (Bramanti et al., 2009), attualmente presente in Europa con una frequenza compresa tra l'1 e il 7%, ma che arriva oltre il 20% nelle popolazioni baltiche e attorno al 40% nei Saami. Inoltre, quasi tutti i

cacciatori-raccoglitori del centro e del nord-est Europa sequenziati sino ad oggi e la maggior parte dei cacciatori-raccoglitori post-neolitici, presentano linee mitocondriali associate all'haplogrupo U (Bramanti et al., 2009; Malmstrom et al., 2009) (Fig 1.4a,c). Ci sono solo tre eccezioni: due individui italiani e uno scandinavo con aplogruppo N\* e linee pre-HV (Caramelli et al., 2003; Skoglund et al., 2012). In tutti gli altri campioni di cacciatori-raccoglitori le linee mitocondriali oggi più diffuse – H, T, K e J – sono assenti, suggerendo che questi aplogruppi sono stati introdotti in Europa nel corso della neolitizzazione.

Un cambiamento del profilo genetico degli Europei durante la transizione neolitica è emerso anche da un recente studio condotto sul genoma di tre cacciatori-raccoglitori e un agricoltore scandinavo, neolitici (Skoglund et al., 2012). L'agricoltore risulta geneticamente molto più simile agli attuali abitanti del sud Europa, in netto contrasto con i cacciatori-raccoglitori che invece risultano più simili agli Europei del nord.







**Fig. 1.4:** mappe europee suddivise per epoche che mostrano le regioni dalle quali provengono le sequenze mitocondriali antiche a disposizione. I quadrati indicano i cacciatori-raccoglitori e i cerchi rappresentano gli agricoltori. Le linee mitocondriali U sono rappresentate in rosso, le altre in giallo. (a) Cacciatori-raccoglitori paleolitici e mesolitici (sulla mappa è rappresentata l'Europa durante l'LGM circa 22.000 BP). (b) Primi agricoltori 7.600-6.500 BP. La mappa illustra approssimativamente il tempo di arrivo e la durata delle più antiche culture neolitiche. (c) Ultimi cacciatori-raccoglitori 6.500-4.500 BP. Mentre i più antichi cacciatori-raccoglitori portano esclusivamente l'hg U, gli ultimi mostrano linee mitocondriali aggiuntive presenti nei primi gruppi neolitici. (d) Tardi agricoltori 6.500-4.500 BP. Rispetto al primo periodo di comparsa, le successive popolazioni agricole mostrano un significativo aumento della frequenza dell'hg U, spiegabile con un aumento del tasso di admixture tra agricoltori e cacciatori-raccoglitori, con conseguente adozione del nuovo stile di vita da parte di questi ultimi. (Immagine da Pinhasi et al., 2012).

Questi risultati suggeriscono che la diffusione dell'agricoltura ha seguito una gradiente sud-nord e che la mescolanza sulla scia di questa espansione ha modellato l'assetto genomico dell'Europa moderna. Nonostante questo studio sia stato condotto su un campione limitato, è comunque coerente con altri lavori basati sull'adDNA, nell'indicare una differenza genetica tra gli ultimi cacciatori-raccoglitori e i primi agricoltori.

Ad oggi, tutti gli studi di paleogenetica sono concordi nel descrivere uno scenario in cui gli agricoltori colonizzarono l'Europa partendo da sud/sud-est e che una frazione consistente del patrimonio genetico mesolitico è sopravvissuta solo nelle popolazioni del nord Europa. Questo modello è supportato anche dai

risultati provenienti da altre fonti: le analisi craniometriche indicano che la affinità tra le prime popolazioni neolitiche dell'Europa centrale con quelle del Vicino Oriente si spiegano grazie al modello della diffusione demica, mentre nel nord Europa i cacciatori-raccoglitori avrebbero adottato gli elementi culturali neolitici dalle vicine comunità agricole senza un importante contributo genetico da parte di queste ultime (Pinhasi & von Cramon-Taubadel, 2009; von Cramon-Taubadel & Pinhasi, 2011; Pinhasi & von Cramon-Taubadel, 2012). In aggiunta, gli studi archeozoologici sull'origine e la diffusione delle principali specie domestiche – maiale, capra, pecora e bovino – hanno aperto nuove prospettive sulla dispersione dell'agricoltura così come sui processi di domesticazione. Essi indicano che tutte e quattro le specie sono state domesticate per la prima volta nella regione a sud-est dell'Anatolia, della Siria e dell'Iraq e nell'Iran occidentale attorno ai 10.500 anni fa e successivamente introdotte in Europa dagli agricoltori in arrivo dall'Anatolia (Vigne, 2011; Zeder, 2011). Gli studi paleogenetici sono in accordo con questa visione (Naderi, et al., 2008; Burger & Thomas, 2011; Bollongino et al., 2012), anche se nel caso del maiale essi indicano degli eventi di introgressione dal cinghiale europeo in fase post-neolitica (Larson, 2011).

#### **1.3.4 Agricoltori neolitici**

Oggi sono numerosi gli studi che forniscono dati sull'mtDNA e sul cromosoma Y di individui neolitici. In uno studio del 2005 sono state confrontate le sequenze mitocondriali di 24 campioni neolitici associati alla cultura della ceramica lineare (Linearbandkeramik, LBK) con 480 Europei moderni provenienti dalla medesima regione geografica ed è emerso che il 25% dei campioni antichi presentano l'aplogruppo N1a, presente solo nello 0.2% dei moderni (Pavúk, 2005; Deguilloux et al., 2011). In uno studio successivo (Bramanti et al., 2009) sono state confrontate le sequenze di 20 cacciatori-raccoglitori per un range temporale compreso tra 15.500 e 2.300 cal. BP, con quelle di 25 agricoltori del periodo LBK e con sequenze moderne: è emerso che tra i tre gruppi esistono profonde differenze genetiche che non possono certamente essere spiegate con la sola continuità tra le popolazioni. Poiché la differenziazione temporale passa attraverso fenomeni di deriva genetica, soprattutto nelle popolazioni di piccole dimensioni, i test sulle ipotesi di "population discontinuity" sono stati svolti considerando le simulazioni di coalescenza unite alla storia delle variazioni nelle dimensioni della popolazione. Tali simulazioni hanno escluso fenomeni di continuità, se ne deduce pertanto che gli agricoltori della cultura LBK del centro Europa non discendono dai cacciatori-raccoglitori del post-LGM, ma verosimilmente sono giunti attorno ai 7.5 ka (Fig. 1.4b) dal nord dei Balcani (Pavúk, 2005). Questi risultati sono supportati anche da studi paleogenetici condotti sulle specie bovine domestiche,

che individuano nelle medesime aree geografiche il centro d'origine della cultura casearia (Burger & Thomas, 2011), a cui si associa anche una coevoluzioni genetica attraverso la persistenza della lattasi (Itan et al., 2009).

Attualmente ci sono pochi altri studi, oltre a quelli sopra descritti, che confrontano le popolazioni mesolitiche e neolitiche. Al contrario sono numerosi i lavori che si sono occupati delle popolazioni neolitiche e dell'Età del Bronzo di diverse località europee: i modelli che emergono non sono omogenei, ma si può supporre che tale disomogeneità sia il frutto di variazioni regionali nelle modalità di diffusione dell'agricoltura. Per esempio, in un recente studio sono stati caratterizzati geneticamente per l'mtDNA 364 individui per un arco temporale che va dal Neolitico all'Età del Bronzo (da 5.500 a 1.550 anni fa) con lo scopo di studiare la variazione cronologica della composizione genetica in Europa centrale. Il confronto con 198 mtDNA di individui mesolitici, neolitici e dell'Età del Bronzo provenienti da diverse parti dell'Europa e con un database di 67.996 sequenze attuali eurasiatiche, ha evidenziato quattro cambiamenti significati nella composizione genetica durante tutto il Neolitico, rivelando un ruolo chiave delle culture tardo neolitiche nella formazione della moderna diversità genetica dell'Europa centrale (Brandt et al., 2013). Hervella et al. hanno confrontato le sequenze mitocondriali di 49 individui provenienti dal nord della Spagna, cronologicamente distribuiti dal Paleolitico Superiore all'Età del Bronzo, con altre sequenze europee preistoriche (Hervella et al., 2012). Le distanze di  $F_{st}$  e i multidimensional scaling indicano che non sono presenti differenze statisticamente significative tra i cacciatori-raccoglitori iberici del centro Europa e della Scandinavia; tuttavia i campioni neolitici spagnoli differiscono dai campioni LBK e moderni dell'Europa centrale, suggerendo un certo grado di eterogeneità tra popolazioni neolitiche di regioni diverse, a supporto di un modello di dispersione random per la diffusione del Neolitico in Europa. Inoltre, le differenze tra sette sequenze mitocondriali neolitiche e sei aplotipi Y-cromosomici provenienti dalla grotta di Avellaner al nord della Spagna (Lacan et al., 2011) e campioni neolitici del centro Europa, sono interpretate come la testimonianza di rotte di espansione diverse lungo il bacino mediterraneo per la diffusione del Neolitico in Europa. L'assenza di una continuità diretta tra i primi agricoltori iberici e gli attuali spagnoli (Gamba et al., 2012), così come tra gli ultimi cacciatori-raccoglitori della Svezia e le moderne popolazioni scandinave (Malmstrom et al., 2009), indicano un turnover popolazionistico diffuso durante l'Olocene. Una conclusione supportata anche dal genoma di cacciatori-raccoglitori e agricoltori scandinavi (Skoglund et al., 2012) e dalle caratteristiche genetiche di Ötzi, affini a quelle dei moderni europei (Ermini et al., 2008; Keller et al., 2012). Al contrario, tutti dati sull'mtDNA nonché i pochi genomi completi oggi a disposizione, evidenziano una forte similarità genetica tra tutti i cacciatori-raccoglitori neolitici e pre-neolitici (Bramanti et al., 2009; Malmstrom et al., 2009; Skoglund et al., 2012).



In conclusione, in Europa i dati che emergono dalla discontinuità tra Neolitico e popolazioni attuali sollevano interrogativi circa i modelli demografici che hanno plasmato la variabilità genetica europea durante e dopo la transizione neolitica. E' ragionevole supporre che in questa fase abbia giocato un ruolo chiave sul moderno assetto genetico europeo, la commistione tra gli agricoltori in arrivo da est e le comunità locali di cacciatori-raccoglitori. Tuttavia, rimangono ancora delle incertezze sull'entità e la natura di questi processi sia da un punto di vista diacronico che sincronico: risulta quindi indispensabile implementare le informazioni incrementando, oltre alle analisi genetiche sui reperti fossili umani, le analisi da altre fonti come la morfometria, lo studio delle specie animali domestiche e le datazioni al radiocarbonio.

### **1.3.5 Espansioni post-neolitiche**

Abbiamo visto come gli studi genetici hanno consentito di individuare le diverse componenti della struttura genetica delle moderne popolazioni europee, indentificando una componente paleolitica africana, una post-LGM e una neolitica dal Vicino Oriente. Tuttavia, altre espansioni successive alla diffusione neolitica hanno lasciato una traccia nel panorama genetico europeo, alcune delle quali ben documentate da un punto di vista archeologico. Ad esempio, circa 5.000 anni fa nelle steppe al sud della Russia si sviluppò una nuova cultura – definita Kurgan – caratterizzata dalla domesticazione del cavallo, che ha dominato l'intera steppa per poi diffondersi ad est, nel nord e nel centro Europa tra i 4.400 e i 2.800 BP: Marija Gimbutas associa questa espansione culturale con la diffusione della lingua indoeuropea (Gimbutas, 1970). Anche durante l'Età dei Metalli, la diffusione di particolari culture materiali e religiose potrebbero essere associate a movimenti popolazionistici paralleli: è il caso per esempio della così detta "cultura dei campi di urne" (urnenfelder) individuata dalla sostituzione dell'inumazione con la cremazione, e della sepoltura dei resti cremati in urne. Questa cultura si sviluppa nell'Europa centrale tra il XIII e metà dell'VIII secolo a.C. per poi diffondersi dal centro Europa sino all'Ungheria, alla Francia occidentale e dalle Alpi sino alle coste del mare del Nord.

Alcuni autori considerano pressoché nullo l'impatto di queste ultime espansioni sull'assetto genetico delle popolazioni europee, mentre altri attribuiscono loro un effetto minoritario, come nel caso di certi studi di frequenza allelica sui marcatori genetici classici (Menozzi et al., 1978; Cavalli-Sforza et al., 1993, 1994; Piazza et al., 1995). Secondo Cavalli-Sforza et al. queste espansioni non hanno fornito un contributo sostanziale alla struttura genetica europea, dal momento che la densità della popolazione durante l'Età del Bronzo era più alta rispetto al Neolitico.

#### 1.4 IL CASO DELL'ITALIA

L'Italia, grazie alle sue particolari caratteristiche geografiche e climatiche, ha occupato sin dalla preistoria una posizione centrale nel flusso di scambi tra antiche popolazioni. Le numerose espansioni, dispersioni e immigrazioni hanno lasciato una traccia indelebile nel complesso mosaico sia culturale che genetico dell'Italia attuale. A partire dall'arrivo dei primi "italiani", appartenenti ad altre specie del genere *Homo* che vissero nella penisola prima di *H. sapiens*, l'Italia è sempre stata terreno di incontro e di transiti. La più antica presenza umana è attestata a quasi un milione di anni fa – Monte Poggiolo circa 800.000 anni fa e Isernia la Pineta almeno 700.000, che hanno restituito industrie litiche di modo 2 – benché i primi ritrovamenti fossili siano ben più recenti. Degno di nota è sicuramente il reperto noto come "uomo di Ceprano", un cranio privo di faccia rinvenuto nella cittadina laziale di Ceprano e recentemente ridatato a 400.000 anni fa (Manzi et al., 2010). Si tratta di un reperto unico nel suo genere, dal momento che le analisi morfologiche indicano che potrebbe rappresentare una variante arcaica di *H. heidelbergensis* (Mounier et al., 2011), da cui presero avvio le distinte storie evolutive dei Neanderthal in Europa e della nostra specie in Africa.

Successivamente, per decine di millenni, l'Italia è stata parte del territorio dei Neanderthal, che al tramonto della loro storia evolutiva hanno convissuto con *H. sapiens*. Dai siti pugliesi a quelli laziali, dalle Prealpi del veronese alla Liguria: a partire da 40.000 anni fa la penisola era abitata da due forme umane vicine, con qualche buona possibilità d'incontro tra le due, come è emerso da uno studio genetico/morfologico condotto su una mandibola neandertaliana proveniente dai Monti Lessini, nel veronese (Condemi et al., 2013), al contrario di quanto risultava da analisi precedenti, che mostravano, a livello mitocondriale, una chiara discontinuità genetica tra le due forme (Caramelli et al., 2006, 2008).

Durante l'LGM la nostra penisola è stata sottoposta all'ondata migratoria da parte dei gruppi umani provenienti dall'Europa centrale che, per far fronte al deterioramento climatico, rapidamente si ritirarono nei rifugi glaciali sulle coste del Mediterraneo - come testimoniano i casi delle penisole iberica e dei Balcani (Achilli et al., 2004; Rootsi et al., 2004; Pereira et al., 2005; Soares et al., 2010) – per poi ri-colonizzare il centro e il nord Europa tra i 16 e i 13 ka (Richards et al., 2000; Rootsi et al., 2004; Semino et al., 2000; Pala et al., 2009, 2012). In seguito, ha visto il passaggio di diversi gruppi umani, sia in epoca preistorica che storica, agendo come punto di arrivo e/o zona di espansione durante i diversi eventi migratori principali successivi alla prima colonizzazione paleolitica.

### **1.4.1 Neolitizzazione della penisola italiana**

Le sponde del Mediterraneo sono interessate da una rapida diffusione della cultura neolitica, come testimoniano la comparsa delle specie domestiche e della ceramica impressa. In Italia la transizione neolitica è ben documentata da un punto di vista archeologico e le più recenti sintesi cronologiche (Passina & Tinè, 2008) descrivono le prime fasi della neolitizzazione come il risultato di due processi indipendenti e paralleli, che coinvolgono, rispettivamente, l'Adriatico – dove sono riconoscibili aspetti della Ceramica Impressa datati a circa 8.100 anni fa (Puglia e Basilicata, Italia meridionale-orientale) – e le coste tirreniche dove si attesta la presenza della Ceramica Impressa Ligure attorno ai 7.900 anni fa (Liguria, Italia nord-occidentale). Questi orizzonti culturali costituiranno la matrice di diverse piccole realtà locali, provviste dell'intero pacchetto neolitico (agricoltura, ceramica e altri elementi) e disseminate lungo l'intera penisola.

Come per il resto dell'Europa, anche per l'Italia restano ancora alcuni interrogativi su quali siano state le modalità di diffusione della nuova cultura. In generale, da un punto di vista archeologico, il Primo Neolitico compare già pienamente sviluppato e lo studio delle tappe e dei meccanismi di neolitizzazione risente delle scarse conoscenze disponibili per il Mesolitico, in particolare per la sua fase più recente: risulta quindi difficile stabilire se sia avvenuta una graduale transizione dei cacciatori-raccoglitori, una repentina sostituzione delle comunità locali da parte degli agricoltori neolitici in arrivo dal Vicino Oriente, o un mix di questi due eventi. I reperti a disposizione mettono comunque in evidenza una graduale transizione al sud della penisola, dove le culture neolitiche sembrano inizialmente innestarsi su industrie mesolitiche, mentre nelle zone settentrionali la comparsa del Neolitico sembra essere più repentina.

Il Neolitico finale e le successive Età dei Metalli rivelano un complesso mosaico di diverse origini culturali, potenzialmente associate alle migrazioni dei diversi gruppi umani. Durante il primo millennio a.C., l'Italia ha ospitato una vasta serie di differenti popoli le cui origini in molti casi restano ancora dibattute (per esempio: Etruschi, Liguri, Veneti), mentre in altri casi sono il risultato di noti eventi migrazionali (i Celti in Italia nord-occidentale, i Greci in Italia meridionale e in Sicilia) (Pesando, 2005). Inoltre, successivi eventi storici legati al commercio e all'espansione delle diverse popolazioni più recenti, come fenici, greci, cartaginesi, romani, arabi e barbari, hanno contribuito all'attuale composizione genetica dell'Italia. Difficile credere che tali eventi possano aver cancellato le tracce genetiche precedenti: più verosimilmente queste migrazioni hanno determinato modelli di diffusione parzialmente sovrapposti ai più antichi, determinando una composizione del pool genico attuale tra le più peculiari del bacino mediterraneo.

### 1.4.2 Le origini del patrimonio genetico italiano

Ad oggi sono pochi gli studi che si sono occupati di ricostruire la storia e la struttura genetica del popolo italiano. Tra questi, Barbujani e colleghi in uno studio del 1995 basato sulla variabilità dell'mtDNA, hanno identificato un gradiente nord-sud con una variabilità genetica intrapopolazionistica maggiore rispetto a quella interpopolazionistica (Barbujani et al., 1995), confermando quello che era emerso precedentemente da uno studio sui marcatori genetici classici (Cavalli-Sforza et al., 1994), mentre sottolineano una sostanziale differenza genetica tra la Sardegna e la penisola. Studi più recenti si sono focalizzati su specifiche regioni italiane e hanno rivelato patterns di distribuzione degli aplogruppi piuttosto omogenei. Questi risultati portano a concludere l'esistenza di una certa uniformità del patrimonio genetico mitocondriale all'interno delle diverse aree dell'Italia (Turchi et al., 2008; Ottoni et al., 2009). Della prospettiva patrilineare si sono occupati Di Giacomo e collaboratori nel 2003, analizzando la diversità del cromosoma Y lungo l'Italia continentale: anch'essi evidenziano un gradiente nord-sud, mentre la distribuzione osservata della variabilità genetica è associata a fenomeni locali di deriva genica e effetto del fondatore. Lo studio è stato replicato da Capelli et al. incrementando il numero di marcatori genetici e adottando una strategia di campionamento più specifica. Essi hanno rilevato che il 70% della variabilità osservata è distribuita lungo un gradiente latitudinale e che le differenze tra le regioni settentrionali e quelle meridionali sono correlate ai diversi contributi neolitici/mesolitici nelle due aree, in accordo con i dati archeologici (Capelli et al., 2007). Questi risultati sono stati confermati da un recente update del medesimo lavoro, nel quale è stato aumentato il campione e i dati sul cromosoma Y sono stati integrati con quelli dell'mtDNA (Brisighelli et al., 2012). La discontinuità tra nord e sud Italia era comunque già emersa da studi condotti a livello genomico sia su scala europea (Lao et al., 2008; Nelis et al., 2009) sia su specifici campioni italiani (Di Gaetano et al., 2012) dai quali emerge sempre la particolare distribuzione della variabilità genetica e la sotto-strutturazione della popolazione italiana.

I recenti dati sulla neolitizzazione dell'Europa meridionale (Lucan et al., 2011a, 2011b) suggeriscono un'altra possibile interpretazione della distribuzione della variabilità genetica, proponendo un modello sex-biased delle migrazioni neolitiche, che avrebbe prodotto l'attuale struttura asimmetrica dell'Italia. Le più recenti migrazioni avrebbero poi incrementato questa asimmetria, come dimostra l'arrivo dei primi gruppi greci nell'Italia meridionale e in Sicilia, sbilanciato verso un basso numero di femmine (Pesando, 2005). Un recente studio condotto sui marcatori uniparentali di circa 900 individui distribuiti lungo tutto il territorio italiano (Sicilia e Sardegna comprese) hanno evidenziato dei patterns sesso-specifici, testimoni di storie demografiche diverse per i maschi e le femmine (Boattini et al., 2013). In oltre, escludendo il caso della Sardegna, anche questo studio evidenzia l'esistenza di un gradiente nord-ovest/sud-est del cromosoma Y, confermando ulteriormente

l'ipotesi di due processi indipendenti e paralleli di neolitizzazione della nostra penisola. Contemporaneamente, una distribuzione più omogenea della diversità mitocondriale è in accordo con gli eventi demografici più antichi, quando l'Italia fu utilizzata come rifugio glaciale durante l'LGM (Boattini et al., 2013).

Ancora una volta tutti questi studi partono da una descrizione della variabilità genetica in campioni contemporanei, dai quali le inferenze vengono teorizzate sul peso relativo della selezione naturale, delle mutazioni, della deriva, delle migrazioni a lungo raggio e del flusso genico tra popoli vicini. Molto spesso inoltre, gli antichi abitanti di una regione sono considerati ancestrali, dunque geneticamente correlati, agli abitanti attuali, senza testare empiricamente questo assunto. Lo studio del patrimonio genetico delle popolazioni antiche risulta quindi indispensabile per ottenere un quadro più dettagliato della struttura genetica della nostra penisola. Ad esempio, in un lavoro del 2010 è stata studiata la variabilità dell'mtDNA dei sardi lungo un arco temporale di 2.500 anni, confrontando 23 sequenze nuragiche dell'HVR1 con 254 individui moderni provenienti dalle regioni dell'Ogliastra e della Gallura (Ghirotto et al., 2010). Le analisi bayesiane hanno restituito uno scenario in cui emerge una certa continuità genealogica tra i Nuragici e gli attuali abitanti dell'Ogliastra ma non con quelli della Gallura, dove la diversità genetica pare essere il risultato di eventi demografici indipendenti.

Lo studio del'aDNA ha fatto luce anche sul dibattuto caso degli Etruschi: il confronto tra l'HVR1 di 30 Etruschi, 27 individui medievali e 370 Toscani moderni, ha dimostrato come gli Etruschi possono essere considerati ancestrali agli attuali abitanti del Casentino e di Volterra, ma non degli altri Italiani che abitano attualmente la vecchia Etruria (Ghirotto et al., 2013). In aggiunta, considerando 158 individui provenienti da due diverse aree dell'Anatolia, il più antico link genetico tra Toscana e Anatolia si colloca attorno ai 5.000 anni fa, suggerendo un'origine autoctona della cultura etrusca e non il risultato di una migrazione dal Vicino Oriente (Ghirotto et al., 2013; Tassi et al., 2013).

Da questi casi emerge chiaramente come lo studio del DNA antico sia di fondamentale importanza per le indagini filogeografiche e microgeografiche, soprattutto per quelle aree in cui il pool genico è il frutto di numerosi eventi demografici che si sono stratificati sin dalla preistoria, com'è il caso dell'Italia.

## **1.5 GLI ABITANTI DELL'ITALIA ANTICA**

Durante l'Età dei Metalli, prima dell'Impero Romano, l'Italia era abitata da numerose popolazioni culturalmente molto diverse (Fig. 1.5), ma legate economicamente da intensi traffici commerciali. Le domande a cui solitamente si cerca di dare una risposta riguardano soprattutto le origini di questi popoli, il

loro contributo al pool genico attuale e le relazioni esistenti tra loro e gli attuali abitanti della medesima aree geografiche. Come descritto nei paragrafi precedenti, l'aDNA rappresenta uno strumento fondamentale per far chiarezza su questi argomenti: lo hanno dimostrato gli studi condotti sui Nuragici (Ghirotto et al., 2009) e sugli Etruschi (Ghirotto et al., 2013; Tassi et al., 2013). Il fatto è che sono ancora troppo poche le analisi condotte secondo questo approccio, e sappiamo ancora troppo poco di numerose popolazioni tardo-neolitiche o post-neolitiche come gli Umbri Plestini, i Messapi, i Piceni ecc...



Fig. 1.5: antichi popoli italici.

### 1.5.1 Gli Umbri Plestini

L'Umbria ha rappresentato fin dalla preistoria un punto nodale per la comunicazione tra il mar Tirreno e il mare Adriatico. In particolare gli altopiani di Colfiorito – o altopiani Plestini (Fig. 1.6) - furono occupati stabilmente fin dalla prima metà dell'Età del Ferro (a cavallo tra il X e il IX secolo a.C.) dagli Umbri Plestini, descritti dalle fonti letterarie antiche come un'aristocrazia montanara simile a quella degli Etruschi.



Fig. 1.6: il territorio degli Umbri Plestini: in evidenza gli altopiani di Colfiorito, o altopiani Plestini

Essi abitarono dapprima sulle sponde dell'antico lago Plestino, ora piano di Casone, in villaggi di capanne, abbandonati nella seconda metà del VII secolo a.C. a seguito di un mutamento climatico che determinò l'innalzamento del livello del lago con conseguenti alluvioni degli abitati. Furono quindi occupate alcune delle alture circostanti, mentre sulle sponde del lago rimase il santuario dedicato alla dea Cupra, la principale divinità femminile del pantheon umbro. Agli inizi del VI secolo a.C. si riscontrò un significativo aumento demografico e i Plestini arrivarono ad occupare un territorio di 120-130 km<sup>2</sup>. L'abitato principale dell'area Plestina appare essere stato quello identificato sulla sommità di Monte Orve, situato quasi al centro dell'altipiano di Colfiorito, che assunse fin dalla fine del VI secolo a.C. caratteristiche urbane (Bonomi Ponzi, 1985). Ai piedi di Monte Orve, sul versante sudorientale, si estende una

vasta necropoli costituita da 190 sepolture in fossa terragna che rappresentano la principale fonte archeologica di informazioni sulla cultura dei Plestini, sulla loro economia, sulla loro organizzazione politica e sociale. Le sepolture scavate fino ad ora sono databili dalla fine del X secolo a.C. agli inizi dell'età romana imperiale.

Per quanto riguarda l'esame dei corredi funerari, si va dalla semplicità di quelli della prima Età del Ferro composti da un solo vaso ad impasto grezzo e da pochi oggetti ornamentali di bronzo, alla ricchezza quasi principesca dei corredi di età orientalizzante (VII secolo a.C.) ed arcaica (VI-V secolo a.C.). Questo ha permesso di ricostruire l'evoluzione della società e della cultura plestina (Bradley, 2000) e dei complessi rapporti di interscambio con le popolazioni confinanti e con quelle più lontane, come gli Etruschi e i Piceni prima, e i Romani poi.

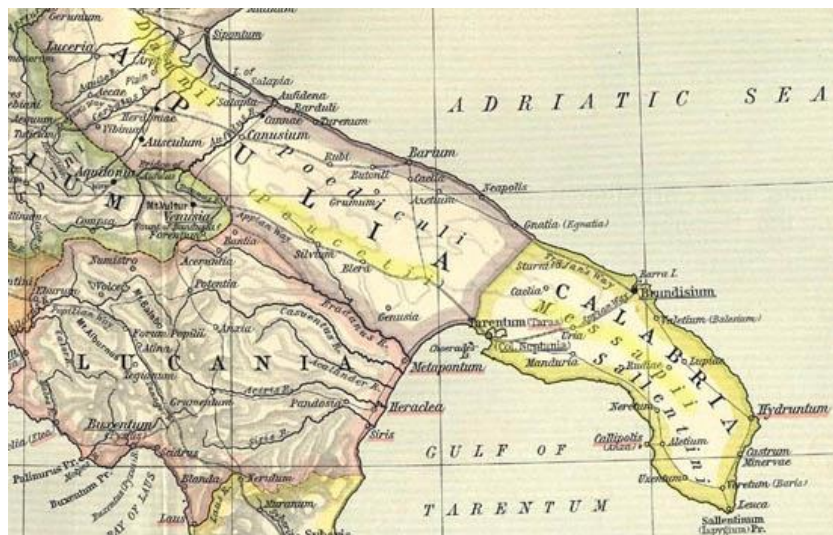
Da tutti gli studi finora condotti sono stati ricavati dati estremamente utili per la ricostruzione della storia del popolamento del territorio italiano in generale e plestino in particolare. Il settore più carente rimane tuttavia quello delle analisi genetiche sui reperti ossei umani, che possono fornire informazioni indispensabili per completare il quadro di questa popolazione.

### **1.5.2 I Messapi**

Durante e dopo l'Età del Ferro la Puglia ospitò una popolazione chiamata Iapigi, un'etnia costituita da Messapi, Peucezi e Dauni stanziatisi nella Messapia, un territorio corrispondente all'attuale Murgia meridionale e al Salento (Fig. 1.7). In particolare, i Messapi sono il popolo che abitò la penisola salentina dal IX secolo a.C. fino al 267/266 a.C. quando, con la fine del *bellum sallentinum*, i Romani assoggettarono la Messapia con il conseguente avvio del cosiddetto processo di romanizzazione.

I Messapi furono un popolo dedito all'agricoltura ed alla pastorizia, riconosciuti anche come abili domatori di cavalli, tenaci combattenti a cavallo ed arcieri. La scarsità delle fonti storiche non permette di conoscere con certezza le origini di questa etnia e degli Iapigi in generale. La prima fonte documentata fu scritta da Esiodo (VIII-VII secolo a.C.): più che di notizie storiche si tratta di tentativi di legittimare le origini degli Iapigi. Infatti il poeta identifica la derivazione del nome Iapigi da Ipeto (figura mitologica greca). Mentre secondo Erodoto (485-425 a.C.) i Messapi provengono dai Cretesi che in seguito ad un naufragio si stanziarono in Puglia prendendo successivamente il nome di Iapigi-Messapi. Alcuni studiosi ritengono che il nome Messapi significhi popolo fra i due mari, altri credono che derivi dal nome del re Messalo. In sostanza le ipotesi storico-archeologiche sulle loro origini sono differenti e discordanti.





**Fig 1.6:** la Puglia antica con i suoi abitanti

I ritrovamenti più antichi associati a questa civiltà sono quelli portati alla luce in scavi archeologici in alcune grotte vicino a Otranto e Roca Vecchia (Pagliara, 1983), mentre i primi insediamenti stabili, datati al VIII secolo a.C., sono stati individuati nelle città di Oria, Cavallino e Vaste (D'Andria, 1989). Attorno al VII-VI secolo a.C. si passa da semplici costruzioni in capanne con zoccolo in pietre irregolari, alzato in mattoni crudi e copertura a rami intrecciati, a costruzioni con più ambienti, di forma quadrangolare, con muretti a secco e copertura a tegole (Van Compernelle, 2001). Quasi tutte le città messapiche erano costruite su un luogo elevato ed erano cinte da una o più cerchie murarie (D'Andria, 1995), costituite solitamente da strutture a blocchi, regolarmente quadrati (Van Compernelle, 2001). Uno dei siti archeologici più ricchi è quello di Muro Tenente presso Mesagne, al confine con il territorio di Latiano, nel sito corrispondente all'antica Scamnium. Si tratta di un insediamento che ha conosciuto varie fasi, dall'Età del Ferro fino al V sec. a.C.; è però la fase messapica – riconducibile con maggiore precisione all'epoca ellenistica (IV-III sec. a.C.) – quella che risulta la più documentata, che ha restituito un sistema formato da una struttura che può definirsi “un edificio residenziale con funzione pubblica” (Cavallo, 1999) e da una attigua necropoli abbastanza estesa. I cumuli di ossa umane erano inumate in tombe a cassettoni e/o in veri e propri ossari, accompagnate da corredi funerari piuttosto ricchi, costituiti da bronzi e brocche dipinte con motivi geometrici.

Se da un lato, grazie alle numerose testimonianze archeologiche, sappiamo definire con precisione la cultura materiale, l'arte e la religione dei

Messapi, lo stesso non può dirsi della loro origine, dei loro rapporti con le altre popolazioni italiche e della loro discendenza. Un'analisi genetica dei reperti ossei umani, mai effettuata sino ad ora, è quindi necessaria per far luce sulla storia del Messapi.

### 1.5.3 I Sardi

Le sue peculiarità geomorfologiche, linguistiche e popolazionistiche fanno della Sardegna una delle regioni più studiate dal punto di vista genetico e demografico. L'isola risulta popolata sin dal Paleolitico, ma è durante la fase neolitica che si manifestano le culture più fiorenti. Tra tutte spicca la civiltà nuragica che deve il suo nome al termine con cui in sardo viene chiamato il monumento considerato più rappresentativo di tale civiltà, il "nuraghe" appunto. Si tratta di un edificio a torre, costruito con l'impiego di pietre di grandi dimensioni (utilizzate grezze o più o meno regolarmente lavorate), al cui interno si trovano una o più camere sovrapposte caratterizzate dalla tipica copertura denominata a "falsa cupola" o "tholos". Intorno ai numerosi nuraghi venivano solitamente edificati i villaggi di capanne in pietra. I dati archeologici consentono di affermare che la civiltà nuragica si reggeva su un'economia agropastorale, ma praticava anche un significativo sfruttamento delle risorse minerarie, in particolare rame e piombo. Dal punto di vista sociale, sembra essere stata caratterizzata da una struttura fortemente gerarchizzata, il cui vertice doveva essere occupato dai guerrieri, ma anche da personaggi legati alle pratiche culturali (Pallottino, 2000). La civiltà nuragica si concentrò prevalentemente in un'ampia fascia che comprende la Sardegna nord-occidentale e centro-occidentale, con esclusione della fascia nord-orientale e centro-orientale e di tutta la parte meridionale dell'isola: si tratta di una distinzione microregionale che tuttavia non coincide con il dato genetico, dal momento che come descritto precedentemente, il confronto delle sequenze mitocondriali nuragiche con quelle dei sardi moderni mostra una certa continuità genealogica tra i Nuragici e gli abitanti dell'Ogliastra (Ghirotto et al., 2009).

I numerosi studi molecolari mostrano come la Sardegna risulti uno dei principali outliers genetici dell'Europa (Cavalli-Sforza & Piazza, 1993). Le sue particolari caratteristiche sono verosimilmente il prodotto evolutivo dell'interazione tra l'isolamento riproduttivo, le ridotte dimensioni della popolazione e la frammentazione degli habitat. A partire da 10 ka l'isola è stata raggiunta da numerose popolazioni diverse (Webster, 1996), ma forti fattori di isolamento, come il mar Mediterraneo e le catene montuose, così come le barriere culturali (Barbujani & Sokal, 1991) hanno probabilmente incrementato gli effetti della deriva genetica, sia all'interno dell'isola che tra essa e il resto dell'Europa. Tuttavia, nonostante i numerosi studi condotti sulle popolazioni

sarde moderne ed antiche, i dettagli di questi processi restano poco chiari, anche perché le analisi sull'aDNA si sono concentrate quasi esclusivamente sulle popolazioni nuragiche (Caramelli et al., 2000; Ghirotto et al., 2009). Risulta quindi indispensabile incrementare il data-set a disposizione, studiando anche le popolazioni sarde neolitiche e pre-nuragiche. Tra i siti che meglio documentano queste epoche preistoriche ritroviamo Cuccuru is Arrius dove è stata rinvenuta una necropoli del Neolitico medio, appartenente alla cultura di Bonuighinu, composta principalmente da tombe scavate a grotticella artificiale, con sepolture per lo più primarie di un solo individuo. Quattro delle circa 20 tombe presenti nella necropoli, sono del tipo a fossa: si tratta delle tombe ipogeiche più antiche di tutto il Mediterraneo.

Altro sito neolitico degno di nota è quello nella località di Su Forru de is Sinzurreddes, nel territorio di Pau. Si tratta di una delle più antiche e importanti stazioni di lavorazione dell'ossidiana sul Monte Arci, frequentata per tremila anni, dal Neolitico antico sino alla prima Età dei Metalli. Gli scavi, diretti dal ricercatore Carlo Lugliè del dipartimento di Scienze Archeologiche dell'Università di Cagliari, hanno restituito un'antica sepoltura risalente al Neolitico medio a conferma che in quel luogo era presente anche un insediamento umano.

Nonostante i siti pre-nuragici siano relativamente abbondanti, soprattutto nella zona sud dell'isola, il record fossile umano associato e questa facies risulta qualitativamente scarso, il che rende complesse le analisi genético-molecolari di questi reperti.

## **1.6 IL DNA ANTICO: 30 ANNI DI STORIA**

La nascita della paleogenetica risale ormai a trenta anni fa, quando un gruppo di ricercatori ottenne la prima sequenza nucleotidica da un organismo estinto (Higuchi et al., 1984): da allora lo studio del DNA antico ha compiuto enormi passi in avanti. La sua potenza risiede nel fatto che è in grado di fornirci una finestra sul passato che il DNA moderno e gli studi paleontologici da soli non possono offrirci. Come descritto nei precedenti paragrafi, è stato ampiamente adottato per affrontare questioni relative alla storia e alle relazioni tra gli ominidi (Green et al., 2010), alla domesticazione di piante e animali (Leonard et al., 2002; Edwards et al., 2007; Larson et al., 2010; Stiller et al., 2010), per studiare le dinamiche popolazionistiche (Leonard et al., 2000; Shapiro et al., 2004; Campos et al., 2010; Pinsky et al., 2010), la filogenesi degli organismi estinti (Shapiro et al., 2002; Krause et al., 2008; Dabney et al., 2013; Orlando et al., 2009, 2013) e per indagare i cambiamenti genetici in alcuni batteri responsabili di patologie che colpiscono l'uomo (Bos et al., 2011;

Schuenemann et al., 2013). Tuttavia, se da un lato l'aDNA rappresenta un potente strumento di indagine, dall'altro deve essere maneggiato con cautela, dal momento che può indurre a gravi errori interpretativi legati al problema delle contaminazioni.

### 1.6.1 Principali difficoltà nello studio dell'aDNA

Lo studio del DNA antico presenta non poche difficoltà dovute principalmente alla sua natura: scarsa quantità e qualità, alto rischio di contaminazioni, eterogeneità dell'estratto. Di seguito sono descritti i danni subiti dalla doppia elica e le problematiche che ne derivano in fase analitica.

a) *Post-mortem damage*: i danni post-mortem - legati al venir meno dei meccanismi di riparazione del DNA dopo il decesso dell'organismo – sono le prime cause di frammentazione e degradazione del DNA recuperato da specie antiche (Tabella 1.1). La conservazione in ambienti freddi può ridurre o inibire l'attività delle nucleasi, riducendo molti dei danni che avvengono immediatamente dopo la morte dell'organismo. Tuttavia, anche i fattori ambientali concorrono nella degradazione del DNA: l'idrolisi porta alla produzione di singoli filamenti frammentando il DNA per scissione diretta o per mezzo della depurinazione, e a lesioni da miscoding – la più frequente è la deaminazione delle citosine in uracile, che causa transizioni C -> T nella sequenza nucleotidica (Lindahl, 1993). L'ossidazione può invece indurre lesioni che bloccano la polimerasi arrestando l'amplificazione o producendo il così detto fenomeno della "jumping PCR" con conseguente produzione di sequenze chimeriche (Pääbo, 1989; Pääbo et al., 1989; Lindahl, 1993). Infine i crosslinks all'interno o tra i filamenti di DNA (Pääbo, 1989), possono bloccare l'attività della polimerasi in fase di amplificazione.

A causa di questi danni, la maggior parte del DNA che sopravvive è solitamente costituito da frammenti corti (inferiori alle 100 bp, base pair) e ricchi di basi alterate (Poinar et al., 2006). L'entità di questi danni è strettamente collegata alla natura del campione e alle sue condizioni di conservazione (Hoss et al., 1996): ambienti freddi, asciutti e con temperature costanti, come le regioni con permafrost o le caverne, sono i luoghi che hanno restituito le specie meglio conservate con bassi livelli di contaminazione, che hanno permesso studi popolazionistici su larga scala, anche prima dell'applicazione delle tecnologie NGS.

DANNO	MECCANISMI	EFFETTI	SOLUZIONI
<b>Frammentazione</b>	Attività delle nucleasi Degradazione microbica Disseccazione, calore... Depurinazione	Sopravvivenza di poco DNA; conservazione di frammenti corti	Amplificazione di frammenti corti (<100-300 bp) sovrapposti
<b>Miscoding lesions per idrolisi</b>	<i>Deaminazioni da miscoding lesions</i> Adenine in ipoxantine Citosine in uracile 5-metilcitosine in timine Guanine in xantine	<i>Misincorporazioni nucleotidiche</i> A -> G C -> T C -> T G -> A	Estrazioni ed amplificazioni multiple; clonaggio; UDG per rimuovere l'uracile
<b>Blocking e Miscoding lesions per ossidazione</b>	<i>Modificazione delle basi</i> 5-OH-5- methylhydantoin (blocking) 5-OH-hydantoin (blocking) 8-oxoguanosine (miscoding G -> T)	No amplificazione "jumping PCR"  Misincorporazioni nucleotidiche	Specifiche polimerasi; clonaggio; amplificazioni multiple
<b>Crosslinks</b>	DNA_DNA crosslink via alchilazione DNA-crosslink con proteine (es: prodotti di Maillard)	No amplificazione	PTB per rimuovere i crosslinks

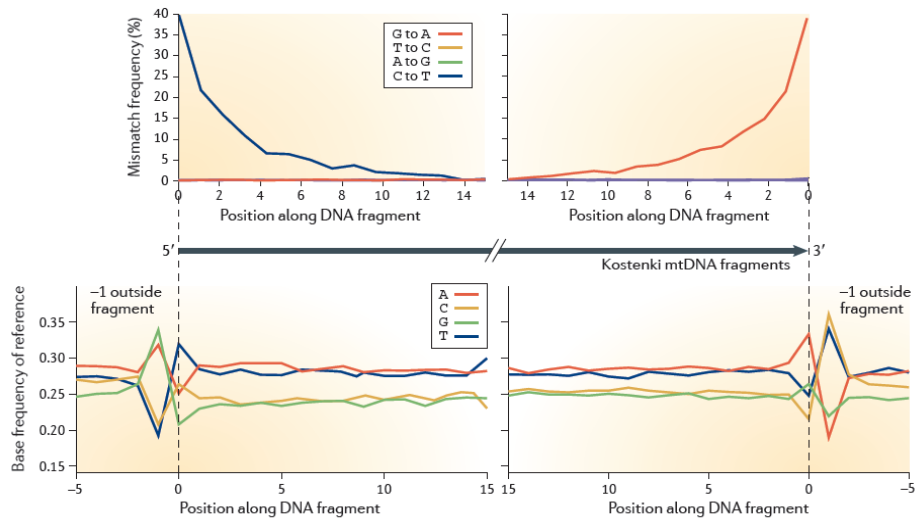
**Tabella 1.1:** danni tipici dell'aDNA

A seguito dei post-mortem damage a cui il DNA va in contro durante il periodo deposizionale, il limite di sopravvivenza della molecola, teoricamente, è compreso tra 100.000 e 1.000.000 di anni, anche se può variare notevolmente in base alle caratteristiche chimico-fisiche degli ambienti di conservazione.

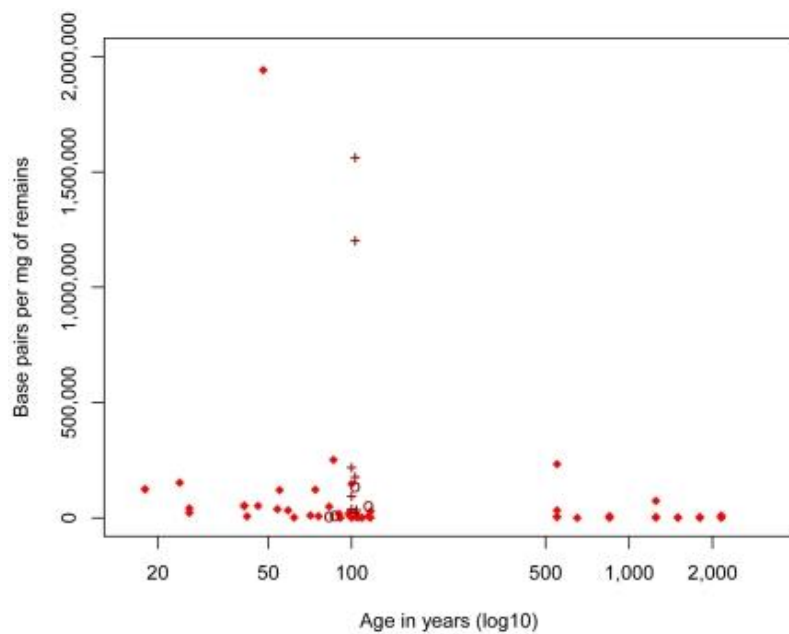
b) Contaminazioni: le difficoltà più ingenti nello studio dell'aDNA nascono dalla presenza di una piccola percentuale di molecole endogene presenti in un estratto, rispetto alla quantità di DNA ambientale. In fase di PCR i contaminanti moderni vengono amplificati preferenzialmente a causa della loro migliore conservazione: se durante i primi cicli di PCR vengono amplificate anche solo poche molecole contaminanti, queste diventeranno rapidamente più numerose riducendo sensibilmente la possibilità di recuperare frammenti antichi.

Ulteriori contaminanti possono essere aggiunti a quelli già presenti nel campione in ogni fase delle analisi, dallo scavo sino alla produzione della sequenza nucleotidica. E' stato dimostrato che ci sono alcuni distretti scheletrici più proni alle contaminazioni (Pilli et al., 2013) e che sono da preferire i reperti mai manipolati e studiati da un punto di vista morfologico (Pilli et al., 2013). In fase sperimentale, i punti più critici sono l'estrazione e l'amplificazione: in questi passaggi, il personale addetto ai lavori può infatti introdurre il proprio DNA o altro materiale genetico presente in laboratorio; pure i reagenti stessi possono essere fonte di contaminazione contenendo spesso DNA umano o animale (Leonard et al., 2007); infine, anche la piccola quantità di aerosol che viene liberato ogni volta che un tubo è aperto, può contenere sino a un milione di copie di DNA in un volume di 0.005 µl: questa concentrazione, potenzialmente, è migliaia di volte maggiore delle copie di DNA endogeno presenti in un reperto antico (Willerslev & Cooper, 2005). A questo proposito, se si opera secondo la metodologia classica è bene attenersi rigorosamente alle regole per la prevenzione delle contaminazioni elencate da Cooper e Poinar (Cooper & Poinar, 2000), oltre che lavorare in locali adibiti esclusivamente allo studio del DNA antico (Pääbo et al., 2004) e tipizzare tutti coloro che vengono a contatto con il reperto per poi confrontare le sequenze di queste con quelle dei campioni (Caramelli et al., 2008; Pilli et al., 2013).

Relativamente più semplice risulta invece la discriminazione dei contaminanti moderni se si adottano strategie d'indagine basate sulle NGS - Shotgun o Target Enrichment – grazie alle quali è possibile visualizzare le caratteristiche biochimiche dei frammenti sequenziati. Le molecole antiche presentano infatti un eccesso di misincorporazioni alle estremità dei filamenti prodotte della degradazione; alcuni studi hanno evidenziato che circa il 40% delle citosine presenti alle estremità -5' e -3' sono deaminate, mentre solo il 2% di quelle interne presentano danni post-mortem, verosimilmente perché le porzioni finali dei frammenti sono spesso a singolo filamento (Briggs et al., 2007; Brotherton et al., 2007) (Fig. 1.7a). Tali patterns di sostituzioni nucleotidiche vengono utilizzati per discriminare le molecole endogene dai contaminati moderni e per stimare la lunghezza media dei frammenti in funzione dell'antichità del reperto (Fig. 1.7b) (Briggs et al., 2007; Sawyer et al., 2012), anche se sappiamo che la conservazione del DNA non è direttamente collegata all'antichità del campione.



**Fig. 1.7a:** patterns di misincorporazioni e di frammentazione nel genoma mitocondriale di un sapiens arcaico russo. L'incremento delle sostituzioni nucleotidiche è funzione della distanza dalle estremità -5' e -3' dei filamenti (immagine da Stoneking & Krause, 2011).



**Fig. 1.7b:** dimensioni (bp) dei frammenti endogeni di mtDNA determinati per milligrammo (mg) di tessuto in funzione dell'età del reperto (immagine da Sawyer et al., 2012.)

Nonostante vengano seguiti alla lettera i comportamenti per la prevenzione delle contaminazioni, non è certo che si abbiano risultati autentici: per questa ragione l'onere del ricercatore è quello di analizzare criticamente il disegno sperimentale, per valutare quali criteri sono pertinenti, e soprattutto, determinare se i risultati ottenuti hanno un senso filogenetico (Pääbo et al., 2004; Gilbert et al., 2005).

### 1.6.2 Tecniche di analisi del DNA antico

Sino a pochi anni fa le strategie d'indagine per l'analisi dell'adna a disposizione dei ricercatori si limitavano all'amplificazione mediante PCR delle porzioni genomiche oggetto di studio. Un approccio che senza dubbio ha portato a grandi risultati, ma che nello stesso tempo presenta dei limiti che oggi possono essere superati con l'utilizzo delle metodologie NGS.

a) Metodologia classica: viene utilizzata sin dai primi anni '80; prevede l'amplificazione dei frammenti d'interesse, il clonaggio batterico dei prodotti di amplificazione e il sequenziamento degli ampliconi con metodo di Sanger (Higuchi, 1984; Pääbo, 1985, 1989), con le opportune modifiche ai protocolli standard per renderli adeguati alle problematiche del DNA degradato. Per esempio, dal momento che le dimensioni medie dei frammenti antichi si aggirano attorno alle 100 bp, quando si vogliono amplificare frammenti lunghi, la regione target viene suddivisa in più porzioni parzialmente sovrapponibili e amplificata per mezzo di più coppie di primers: la sequenza completa viene poi ricostruita in fase di analisi. I prodotti di PCR sono quindi sottoposti a clonaggio batterico, indispensabile per discriminare gli errori introdotti dalla polimerasi in presenza di siti abasici e/o danneggiati, dalle mutazioni proprie del campione e, infine, ogni singolo amplicone viene sequenziato grazie a dei terminatori marcati a fluorescenza introdotti durante la reazione di sequenza.

Le analisi svolte seguendo questa procedura si concentrano esclusivamente su piccole porzioni genomiche, come la regione di controllo dell'mtDNA o singoli geni nucleari evolutivamente significativi. Nonostante ciò, la "metodologia classica" rappresenta uno strumento potente per indagare la composizione genetica di organismi e popolazioni estinti, e può essere utilizzata per ottenere dati dai quali dedurre inferenze per le relazioni filogenetiche, stimare il tempo di divergenza, la struttura, i patterns filogeografici e di ibridazione delle popolazioni (Caramelli et al., 2007; Mona et al., 2010; Ghirotto et al., 2009, 2010; Guimaraes et al., 2009; Hughey et al., 2013; Brandt et al., 2013), così come i cambiamenti funzionali e gli adattamenti genetici ai diversi ambienti in organismi moderni ed estinti (Lalueza-Fox et al., 2007; Krause et al., 2007; Lari et al., 2010; Wang et al., 2012).



Tuttavia, data la natura e la limitata quantità di materiale fossile, gli approcci tradizionali basati sulla PCR - anche quando eseguita in multiplex - consentono di ricostruire sequenze lunghe al massimo un migliaio di paia di basi (Krause et al., 2006). A questo si aggiunge il fatto che i frammenti più corti che si possono amplificare mediante PCR si aggirano attorno alle 150 bp, poiché sequenze più corte si avvicinano alle dimensioni dei primers: dato che l'aDNA è caratterizzato da frammenti estremamente corti (Poinar et al., 2006; Green et al., 2008, 2009), utilizzando la PCR si riduce la possibilità di recuperare molecole endogene. Per queste ragioni, l'high-throughput sequencing offre diversi vantaggi per l'analisi del DNA antico.

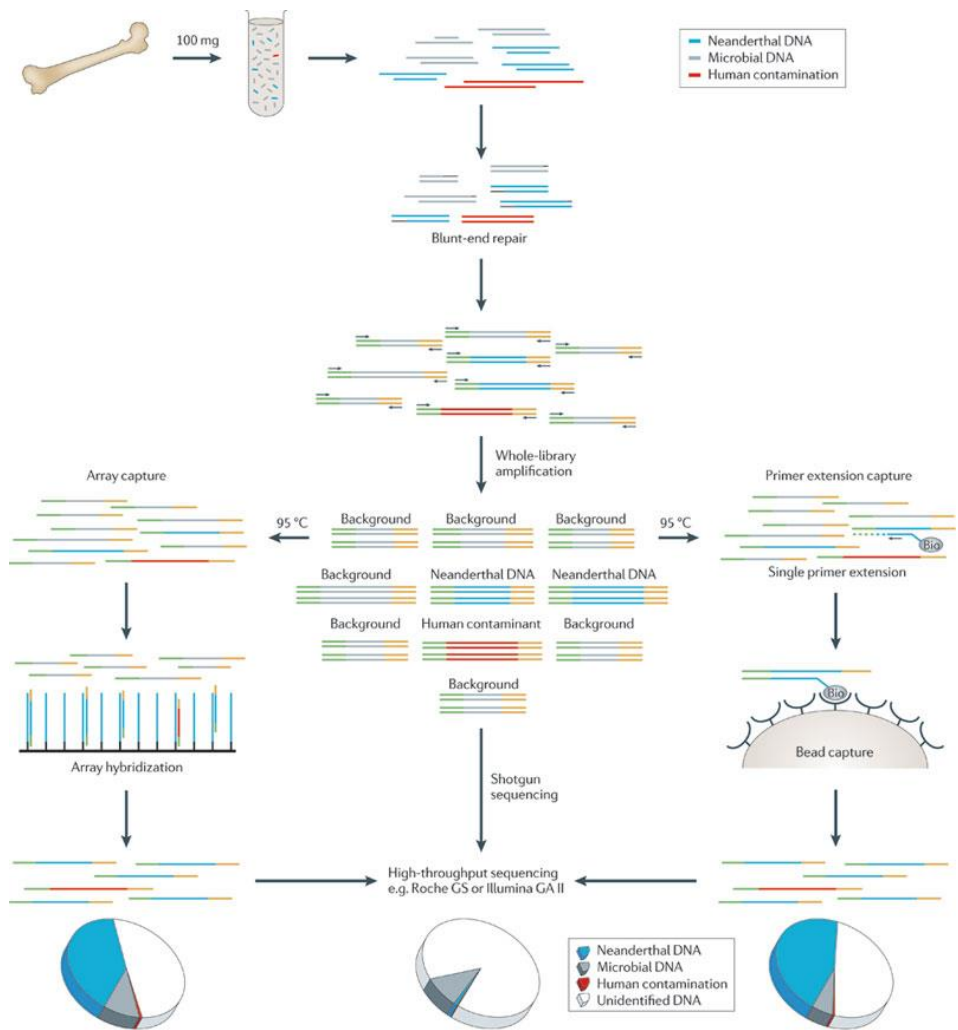
b) *High-throughput sequencing*: grazie all'applicazione della tecnologia NGS allo studio dell'aDNA siamo passati dalla paleogenetica alla paleogenomica, con il conseguente ribaltamento di alcune teorie circa l'origine e l'evoluzione di *H. sapiens* e i rapporti di quest'ultimo con le forme arcaiche con le quali è venuto a contatto una volta uscito dall'Africa (Green et al., 2010; Reich et al., 2011).

Il sequenziamento ultramassivo non richiede l'amplificazione del DNA target mediante PCR, ma l'estratto viene convertito direttamente in libreria attraverso la ligazione di adattatori alle estremità dei frammenti: l'utilizzo di adattatori campione-specifici (MID e INDEX) fa sì che si possano sequenziare contemporaneamente più librerie, mantenendo l'identità di ciascun campione, e identificare potenziali contaminanti; in più le librerie possono essere amplificate via PCR utilizzando primers complementari agli adattatori (Fig.1.8). Anche per il sequenziamento ultramassivo sono stati messi a punto protocolli specifici per il DNA antico (Briggs et al., 2009; Maricic et al., 2010; Gansuage & Meyer, 2013) che consentono di ottenere genomi completi partendo da quantità estremamente ridotte di materiale fossile (Reich et al., 2010). Come descritto precedentemente, il sequenziamento diretto delle librerie consente inoltre di discriminare il DNA endogeno da potenziali contaminanti mediante l'analisi dei patterns di misincorporazione alle estremità dei filamenti (Fig. 1.7a). Infine, ogni frammento, grazie alle sue ridotte dimensioni, può essere completamente sequenziato da entrambe le estremità, il che permette di identificare eventuali errori di sequenziamento (Green et al., 2010).

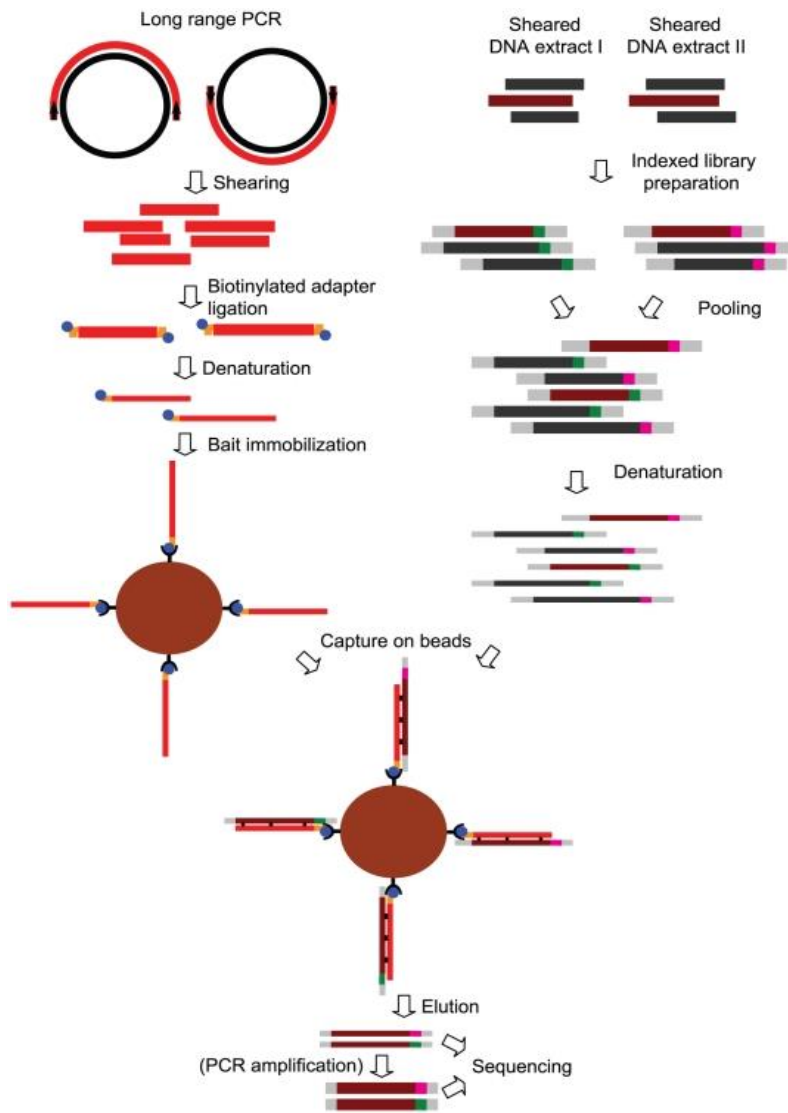
Dall'altro lato, uno dei principali limiti nel sequenziamento diretto delle librerie è legato alla scarsa quantità di DNA endogeno presente nel campione antico, per cui la stragrande maggioranza delle sequenze risultano batteriche, fungine o di altra natura (Fig 1.8). Ovviamente la quantità di materiale informativo che si riesce ad ottenere, varia da campione a campione: ad esempio, il sequenziamento del genoma del mammut è stato completato con 30 milioni di reads partendo da resti che contenevano più del 90% di DNA endogeno (Miller et al., 2008), mentre ci sono volute circa 1.5 miliardi di reads per il genoma completo di Neandertal là dove i campioni si presentavano con meno del 5% di materiale genetico endogeno (Green et al., 2010). Inoltre, l'alta

qualità dei risultati restituiti dai reperti conservati nel permafrost – come appunto il mammut e il Paleo-Eskimo (Rasmussen et al., 2010) – indica, ancora una volta, che la crioconservazione favorisce la sopravvivenza del DNA genomico.

c) *Hybridization capture*: al fine di ridurre i costi di sequenziamento e ottimizzare la resa dei campioni, recentemente sono state messe a punto strategie sperimentali che consentono di selezionare le porzioni genomiche d'interesse (come ad esempio l'mtDNA) mediante sonde di cattura che vengono ibridate al DNA target (Fig. 1.8). Le sonde possono essere sintetizzate in vitro (Burbano et al., 2010) (Fig. 1.8) o prodotte in casa per mezzo di PCR (Maricic et al., 2010) (Fig. 1.9): in ogni caso, là dove c'è complementarità, il DNA si lega alle sonde, mentre quello che resta in soluzione viene eliminato. Alla fine ciò che resta è solo il DNA target arricchito che sarà staccato dalle sonde, eluito e sequenziato. L'hybridization capture può quindi essere applicata anche al DNA antico per superare le limitazioni legate alle piccole percentuali di DNA endogeno: in linea di principio, con questa metodologia è possibile ottenere un genoma completo anche quando il DNA estratto è inferiore all'1%. Inoltre, la stretta relazione evolutiva tra l'uomo e lo scimpanzé, è vantaggiosa allorché le sequenze del genoma dell'uno e/o dell'altro possono essere utilizzate per creare sonde di arricchimento applicabili su qualsiasi ominide estinto (Stoneking & Krause, 2011).



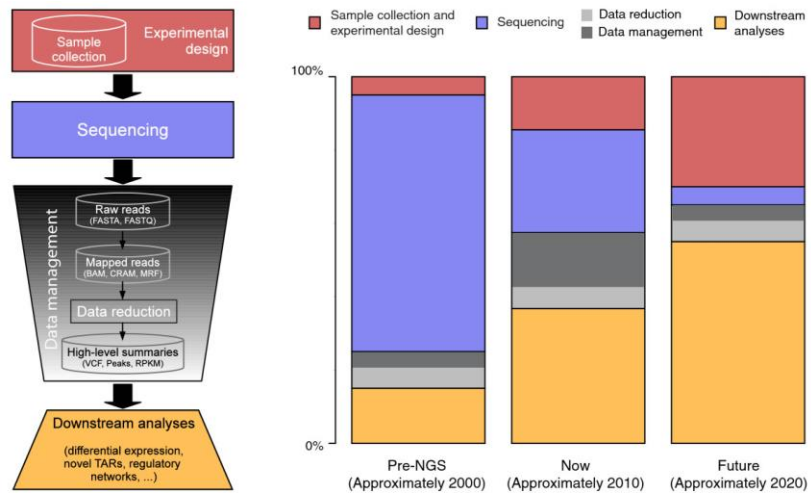
**Fig. 1.8:** varie strategie di sequenziamento ultramassivo per lo studio dell'aDNA (immagine da Stoneking & Krause, 2011).



**Fig. 1.9:** produzione delle sonde di cattura mediante long-range PCR (immagine da Maricic et al., 2010).

d) Assemblaggio delle reads e ricostruzione della sequenza consenso: se da una parte le tecnologie NGS hanno reso possibile sequenziamenti fino a poco tempo fa impensabili, dall'altro lato sono emerse alcune difficoltà legate alla gestione dell'enorme quantità di dati prodotti per ogni run. Si assiste così ad un trend che

vede una continua riduzione nei costi di sequenziamento, accompagnata da un incremento nei costi dell'experimental design, delle analisi e della produzione del dato finale (Sboner et al., 2011) (Fig. 1.10).



**Fig. 1.10:** contributo dei diversi step nel costo complessivo di un progetto di sequenziamento attraverso il tempo (immagine da Sboner et al., 2011).

Nel caso di campioni antichi, in particolare, gli strumenti attualmente disponibili per l'assemblaggio delle reads, presentano importanti limitazioni, a causa della piccola dimensione dei frammenti e della natura ripetitiva di gran parte del genoma dei mammiferi, che crea ambiguità nel mappaggio delle sequenze. Ad esempio, frammenti corti possono causare bias nel mappaggio, oppure non essere mappati accuratamente; o ancora, i frammenti possono mappare in posizioni diverse nei diversi genomi di riferimento a seconda della completezza e accuratezza dei genomi di riferimento stessi (Stoneking & Krause, 2011). Per questi motivi, è richiesto un coverage piuttosto elevato di ogni porzione analizzata al fine di ottenere sequenze consenso di buona qualità, in attesa dello sviluppo di tools specifici per l'assemblaggio di genomi antichi.

Da tutto ciò emerge come la storia evolutiva della nostra specie risulti sempre più definita, grazie alle nuove tecniche di analisi sull'aDNA e ai progressi compiuti sui modelli computazionali per la stima dei parametri

demografici. Tuttavia c'è ancora molto da fare, soprattutto nella costruzione di modelli complessi (quindi più realistici) e comunque, trattandosi della storia umana, ci sono ancora molte storie da scrivere!

## 2. SCOPO DEL LAVORO

L'obiettivo del nostro progetto è stato quello di approfondire le conoscenze circa l'origine dell'attuale struttura genetica della popolazione italiana. In particolare abbiamo focalizzato la nostra indagine sulle espansioni neolitiche e post-neolitiche che hanno interessato le regioni Umbria, Puglia e Sardegna. Per queste zone, infatti, non ci sono ancora sufficienti dati genetici in grado di risolvere alcuni aspetti popolazionistici, come l'origine delle culture presenti in epoche storiche e protostoriche, la natura delle eventuali interazioni tra antiche popolazioni, e infine se e in che misura queste popolazioni hanno contribuito al pool genetico attuale.

Per rispondere a queste domande è stata indagata la variabilità genetica della regione HVR1 del DNA mitocondriale, estratto da reperti umani dell'età post-neolitica provenienti dalle suddette regioni italiane. Il confronto con gli aplotipi mitocondriali delle popolazioni moderne presenti nelle medesime aree geografiche ci ha consentito, quindi, di risolvere alcune questioni microgeografiche e filogeografiche riguardanti la storia della nostra penisola.

## **2. MATERIALI E METODI**

### **3.1 CAMPIONAMENTO DEI REPERTI ANTICHI**

Sono stati recuperati un totale di 123 campioni umani antichi distribuiti tra le regioni Umbria (45), Sardegna (27) e Puglia (51). La prerogativa è stata quella di selezionare i reperti che non sono mai stati sottoposti ad analisi morfologiche e/o morfometriche, così da escludere il rischio di contaminazioni da parte degli archeologi. In più, abbiamo selezionato i distretti scheletrici meno proni alle contaminazioni (Pilli et al., 2013) ed esternamente non troppo danneggiati.

#### **3.1.1 Campioni Umbri**

I resti scheletrici umbri sottoposti ad analisi molecolare provengono dalla necropoli di Colfiorito, che comprende sepolture datate tra il IX e il V secolo a.C.. In quest'area gli scavi sono stati condotti tra il 1970 e il 2006 dalla Soprintendenza per i Beni Archeologici dell'Umbria, sotto la direzione della dott.ssa Laura Bonomi. Nonostante nel corso del tempo siano stati condotti studi approfonditi sulla cultura materiale e sui corredi funerari presenti nelle tombe, al contrario i resti scheletrici umani non sono mai stati eccessivamente manipolati e quindi contaminati con materiale genetico moderno. 45 campioni costituiti in prevalenza da ossa lunghe che si presentavano in ottimo stato di conservazione, sono stati sottoposti all'analisi genetica.

#### **3.1.2 Campioni Pugliesi**

I campioni Messapi provengono da 4 diversi siti archeologici, 3 in provincia di Lecce e 1 in provincia di Brindisi. Si tratta di un totale di 51 campioni datati tra il VI e il III secolo a.C., ad eccezione di quelli provenienti dalla necropoli di Serra Cicora che sono di epoca neolitica.



a) Salve (LE): in questo territorio, gli scavi sono iniziati nel 1987 grazie alla collaborazione tra ricercatori e archeologi dell'Università di Lecce e di Sidney. A seguito delle numerose campagne di scavo è stato possibile ipotizzare una ricostruzione delle fasi principali dell'insediamento: il primo si ebbe attorno al 1.550 a.C. ad opera di abitanti giunti probabilmente da un villaggio vicino ed attratti dalla presenza di una sorgente d'acqua dolce. I campioni utilizzati per le indagini genetiche, appartengono invece alla terza fase abitativa, risalente a circa il 550 a.C. I resti scheletrici, per un totale di 5, si presentano in buono stato di conservazione e tutti sono stati sottoposti ad analisi (Tabella 3.1).

b) Serra Cicora (LE): il sito archeologico di Serra Cicora sorge su un ampio pianoro che si affaccia sulla costa ionica. Sono evidenti due fasi distinte di frequentazione risalenti al VI e al V millennio a.C. In quest'ultima fase il pianoro è stato adibito a necropoli. Gli scavi, condotti dall'Università del Salento, hanno messo in luce svariate evidenze archeologiche, molte delle quali riferibili a manifestazioni culturali e funerarie, e diversi resti scheletrici umani. Al nostro laboratorio sono stati consegnati 12 campioni in buono stato di conservazione, tutti sottoposti ad analisi molecolari (Tabella 3.1).

c) Roca Vecchia (LE): situata lungo la costa adriatica, questa località ha sempre svolto un ruolo di fortificazione a tutto vantaggio dell'entroterra salentino. Oggetto di scavi già negli anni venti del 900, il sito accoglie testimonianze di insediamenti messapici a cominciare dal VI secolo a.C.. L'area sepolcrale da cui provengono i campioni utilizzati in questo studio è datata tra la fine del IV e l'inizio del III secolo a.C.. Per questo sito avevamo a disposizione 20 reperti "vergini" (Tabella 3.1) ancora quasi completamente ricoperti di terra.

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
T. 2A	Femore sx	Serra Cicora/VI-V millennio a.C
T. 3A	Femore dx	
T. 3B	Femore sx	
T. 4A	Femore sx	
T. 5A	Femore dx	
T. 6A	Femore sx	
T. 6B	Femore dx	
T. 6C	Femore sx	
T. 7A	Femore sx	
T. 8A	Femore sx	
T. 9A	Femore dx	
T. 11A	Femore dx	

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
Campione 1-T16	Diafisi ulna sx	Roca Vecchia/IV-III secolo a.C
Campione 2-T4	Falange prossimale	
Campione 3-T4	Diafisi tibia dx	
Campione 4-T5	ND	
Campione5- T6	Diafisi osso lungo	
Campione6-T6	ND	
Campione 7 -T6	Metatarso	
Campione 8-T19	Vertebra cervicale	
Campione 9-T19	Metatarso	
Campione 10-T7	Falange prossimale	
Campione 11-T7	Diafisi radio sx	
Campione 12-T7	Diafisi fibula	
Campione 13-T7	Diafisi fibula	
Camp. 14-T20	Diafisi fibula	
Camp. 15-T20	Diafisi fibula	
Camp. 16-T20	ND	
Camp. 17-T20	ND	
Camp. 18-T20	ND	
Campione 19-T3	ND	
Camp. 20-T4/3	Diafisi femore	

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
Tumulo 7 - 4L7	Omero dx	Salve/2.500-2.300 a.C
Tumulo 7 - 11L7	Omero dx	
Tumulo 7 - 40L5	Omero dx	
Tumulo 7 - 68L7	Omero dx	
Tumulo 7 - 75L6	Omero dx	

**Tabella 3.1:** campioni pugliesi provenienti dalle province di Lecce. In verde sono evidenziati i campioni sottoposti alle analisi genetiche.

d) **Mesagne (BR):** a partire dagli anni '70, il territorio di Mesagne è stato oggetto di indagini da parte della Soprintendenza Archeologica della Puglia. I

numerosi scavi e le indagini scientifiche condotti, hanno portato all'individuazione di diverse necropoli messapiche, in vari quartieri dell'abitato attuale, evidenziando l'importanza di Mesagne quale centro popolato dai Messapi, in posizione strategica tra le due città di Brindisi e Oria. Dalla sua collocazione geografica, l'origine del nome di "Mesagne", che per molti studiosi sta a significare appunto "terra di mezzo".

I campioni utilizzati per le analisi genetiche risalgono all'inizio dell'età ellenistica (IV-III secolo a.C.) (Tabella 3.2), che è l'epoca archeologicamente più rappresentata. I resti scheletrici risultavano piuttosto frammentati, ma nel caso in cui erano disponibili dei denti, abbiamo utilizzato quelli per le analisi del DNA.

CAMPIONE	SITO	SITO/DATAZIONE
S. Pancrazio T1	Dente	Mesagne/IV-III secolo a.C.
S. Pancrazio T2	Dente	
S. Pancrazio T5 236	Frammenti	
S. Pancrazio T6	Dente	
S. Pancrazio T6	Dente	
S. Pancrazio TA	ND	
Duca di Genova D1	Diafisi osso lungo	
Duca di Genova D2	Frammenti	
Duca di Genova T7 237	Dente	
Duca di Genova T5 233	Frammenti	
Duca di Genova T8 231	ND	
Duca di Genova T4 232	Frammenti	
Sandonaci T4/90 235	ND	
Santa Rosa 92/1	Dente	

**Tabella 3.2:** campioni Messapi della provincia di Brindisi. In verde sono evidenziati i campioni sottoposti alle analisi genetiche.

Le analisi genetiche sono state condotte su 31 dei 51 campioni a disposizione (Tabella 3.1 e 3.2). I restanti 20 sono attualmente in fase di studio.

### 3.1.3 Campioni Sardi

I campioni antichi sardi a nostra disposizione per le analisi molecolari provengono da tre diversi siti archeologici della provincia di Oristano e coprono un arco temporale che va dal al Neolitico medio al Neolitico finale. Tutti i reperti sono stati recuperati in campagne di scavo condotte dal 2001 ad oggi, coordinate e dirette dal Dott.re Carlo Lugliè dell'Università degli studi di Cagliari.

a) Su Forru de is Sinzurreddus: si tratta di 123 distretti scheletrici recuperati dalle unità stratigrafiche 1037, 1063, 1064, 1070, 1079 e 1086, appartenenti all'orizzonte Neolitico medio B datato alla seconda metà del V millennio cal. BP. Il contesto delle sepolture era sconvolto dal crollo della sommità della volta che ha prodotto un'ampia frammentazione e dispersione dei resti scheletrici: ad ogni modo dovrebbe trattarsi di una sepoltura composta da 4-5 individui, probabilmente legati da vincolo parentale. I reperti si presentavano in pessimo stato di conservazione ed erano costituiti prevalentemente di corone, gemme dentarie e frammenti di smalto.

Per le analisi popolazionistiche è di fondamentale importanza lo studio di campioni appartenenti a individui diversi: data la difficoltà della ricostruzione morfologica dovuta alla frammentazione, la scelta dei campioni da sottoporre ad analisi molecolare si basa sul riferimento stratigrafico, oltre che sullo stato di conservazione dei reperti. Per questa ragione, i resti analizzati sono stati 6, associati all'US 1070 e 1099 (Tabella 3.3).

b) Necropoli di Cuccuru is Arrius: si tratta di 23 campioni relativi al contesto "Bonu Ighinu" datato al Neolitico medio. Anche per questo sito i campioni risultano estremamente frammentati ma, salvo qualche eccezione, possono essere utilizzati per l'analisi molecolare. Il compromesso tra stato di conservazione dei reperti e scopo del lavoro, ci ha suggerito l'analisi di 3 campioni della tomba T386 e 3 della T388 (Tabella 3.3) e 4 da sepolture singole.

c) Siamaggiore: per questo sito avevamo a disposizione 13 distretti scheletrici associati al Neolitico finale-Eneolitico. Dodici di essi appartengono ad un infante: si tratta prevalentemente di denti decidui e gemme dentarie ma si ha pure una diafisi di osso lungo. Un solo campione è riferito ad un individuo adulto, un premolare inferiore destro con radice ben conservata.

Per le analisi molecolari abbiamo selezionato la diafisi dell'infante e la radice del premolare dell'adulto.

d) Su Carroppu di Sirri: da questo sito ci sono pervenuti 9 campioni datati al VI millennio a.C. Poco frammentati e apparentemente in buono stato di conservazione, sono stati tutti sottoposti ad analisi molecolare.

Dopo aver valutato lo stato di conservazione di tutti i reperti che avevamo a disposizione, abbiamo selezionato 27 campioni che dovrebbero appartenere ad altrettanti individui diversi (Tabella 3.3).

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
SZ91 US 1070	Falange tarso	Su Forru se is Sinzurreddus/Neolitico medio B
SZ92 US 1070	Scapola	
SZ114 US 1070	Dente	
SZ93 US 1099	Falange	
SZ94 US1099	Vertebra	
SZ123 US1099	Dente	

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
CA8 T386	Indeterminato	Cuccuru is arrius/Neolitico medio
CA9 T386	Frammento cranico	
CA13 T386	Frammento di	
CA10 T388	Diafisi femorale	
CA11 T388	Perone	
CA12 T388	Femore	
CA20	Mandibola	
CA21	Dente	
CA22	Dente	
CA23	Mandibola	

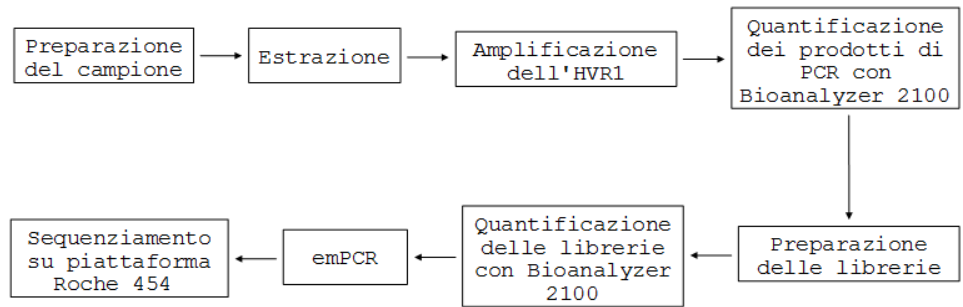
CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
SI01	Diafisi di osso	Siamaggiore/Neolitico finale
SI13	Premolare	

CAMPIONE	DESCRIZIONE	SITO/DATAZIONE
CAR_H010	Mascella con 4	Su Carroppu di Sirri/ Neolitico antico
CAR_H012	Carpale ph1	
CAR_H013	Carpale ph3	
CAR_H014	Diploe cranica	
CAR_H015	Diploe cranica	
CAR_H018	Ulna	
CAR_H019	Tibia	
CAR_H020	Tibia	
CAR_H021	Carpale ph1	

**Tabella 3.3:** campioni sardi selezionati per le analisi genetiche.

### 3.2 METODI

Il marcatore molecolare che abbiamo utilizzato è il DNA mitocondriale, in particolare la regione HVR1 collocata all'interno della D-loop, l'unica porzione non codificante del genoma mitocondriale. L'HVR1, grazie al suo elevato numero di polimorfismi e al suo alto tasso di mutazione, è una regione che si rivela efficace per gli studi filogenetici e popolazionistici di tutte quelle specie, come la nostra, la cui storia evolutiva è recente. L'approccio sperimentale ha visto l'applicazione della metodologia classica combinata al Next Generation Sequencing (Fig. 3.1), che consentono una grande ridondanza di sequenza e un'alta qualità di lettura.



**Fig. 3.1:** workflow per il sequenziamento dei campioni antichi su piattaforma Roche 454.

Tutte le fasi precedenti all'amplificazione sono state svolte in stanze sterili, utilizzate esclusivamente per le analisi sul DNA antico, così da ridurre al minimo la possibilità di contaminazione con DNA moderno.

### 3.2.1 Preparazione del campione

Prima dell'estrazione, ciascun campione è stato sottoposto a pulizia e polverizzazione, entrambe svolte sotto cappa aspirante irradiata prima di ogni utilizzo con raggi UV a 254 nm. La pulizia del campione risulta indispensabile al fine di eliminare, oltre al terreno attorno ai frammenti, eventuali tracce di DNA esogeno. In questo caso è stato rimosso uno strato superficiale di circa 1 mm con carta abrasiva sterilizzata e con micromotore con fresa sterile monouso di 1,8 mm di diametro, come quelli in uso nei laboratori odontoiatrici; ciascuna superficie dei frammenti ossei è stata poi irradiata ortogonalmente per 45 minuti con raggi UV a 254 nm.

Una volta puliti, tutti i frammenti sono stati polverizzati con l'ausilio di un micromotore con fresa sterile monouso di 1,8 mm di diametro: la velocità di rotazione della fresa è stata ridotta al minimo, in modo da evitare il surriscaldamento che potrebbe danneggiare ulteriormente le molecole di DNA. Per ciascun campione sono stati ottenuti dai 150 ai 250 mg di polvere d'osso, conservata alla temperatura di -20°C fino alla fase di estrazione del DNA.

### **3.2.2 Estrazione del DNA**

Il DNA è stato estratto a partire da un'aliquota di polvere d'osso utilizzando un protocollo progettato per massimizzare il recupero di molecole amplificabili e, contemporaneamente, minimizzare la co-estrazione di sostanze che possono inibire l'attività della polimerasi in fase di PCR (Rohland & Hofreiter, 2007).

A circa 250 mg di polvere sono stati aggiunti 5 ml di Extraction Buffer e il tutto incubato overnight a temperatura ambiente. Dopo 2 minuti di centrifugazione a 4.000 rpm, è stato recuperato il surnatante al quale sono stati aggiunti 3,5 ml di Binding Buffer e 100 µl di silice e incubato per 3 ore al buio, a temperatura ambiente. Dopo un altro step di centrifugazione alle stesse condizioni, è stato eliminato il surnatante, mentre il pellet di silice risospeso in 1 ml di Binding Buffer è stato trasferito in una eppendorf da 2 ml e centrifugato al massimo per 10 secondi. Dopo aver rimosso il surnatante, il pellet è stato risospeso in 1 ml di Wash Buffer e centrifugato per 10 secondi al massimo della velocità. Di nuovo è stato rimosso il surnatante e ripetuta la fase di lavaggio; alla fine il pellet è stato fatto asciugare a temperatura ambiente per circa 20 minuti. A questo punto, il DNA è stato eluito in 70 µl di TE 1X per 10 minuti, poi centrifugato per 1 minuto alla velocità massima e il surnatante trasferito in una nuova eppendorf. Infine è stata ripetuta la centrifugazione e il surnatante trasferito nuovamente in una eppendorf pulita, assicurandosi di non lasciare tracce di silice, la quale può sequestrare le molecole di DNA presenti in soluzione. Ogni estratto è stato conservato a -20°C sino alla fase di amplificazione.

Durante l'estrazione è stato portato avanti anche un controllo negativo, nel quale sono stati inseriti tutti i reagenti utilizzati ad eccezione del materiale biologico. Il controllo negativo è stato poi sottoposto ad amplificazione, così da monitorare l'eventuale presenza di contaminazioni nei reagenti utilizzati.

### **3.2.3 Amplificazione dell'HVR1 mediante PCR**

A causa dell'elevata degradazione e frammentazione tipiche del DNA antico, l'HVR1 – lunga 360 bp – non può essere ottenuta in un'unica fase di amplificazione. Per questa ragione sono state utilizzate 3 diverse coppie di primers che amplificano frammenti lunghi circa 170 bp parzialmente sovrapponibili (Tabella 3.4). L'intera regione è stata quindi ricostruita in fase di analisi.



Nome	Sequenza	Lunghezza	Dimensione frammento
L15995	CTACAGGTGGTCAAGTATTTATGGT	25 bp	162 bp
H16132	CCACCATTAGCACCCAAAG	19 bp	
L16107	CGCTATGTATTTTCGTACATTACTGC	25 bp	179 bp
H16261	TGGTATCCTAGTGGGTGAGG	21 bp	
L16247	GATTCACGGAGGATGGT	19 bp	174 bp
H16402	CAACTATCACACATCAACTGCAA	24 bp	

**Tabella 3.4:** coppie di primers utilizzate per l'amplificazione della regione HVR1

Per ciascuna amplificazione sono stati impiegati dai 2 ai 6  $\mu$ l di estratto in una miscela di reazione così costituita: 1X Reaction Buffer (Applied Biosystem), 200  $\mu$ M di ciascun dNTP, 1 mM di ciascun primer e 2 U di Taq polimerasi, in un volume finale di 50  $\mu$ l. La reazione è stata svolta con il seguente profilo termico:

- 95°C per 10 minuti
- ripetizione del seguente ciclo per 50 volte:
  - 94°C per 45 secondi
  - 56°C per 1 minuto
  - 72°C per 1 minuto
- 72°C per 1 minuto
- 4°C forever

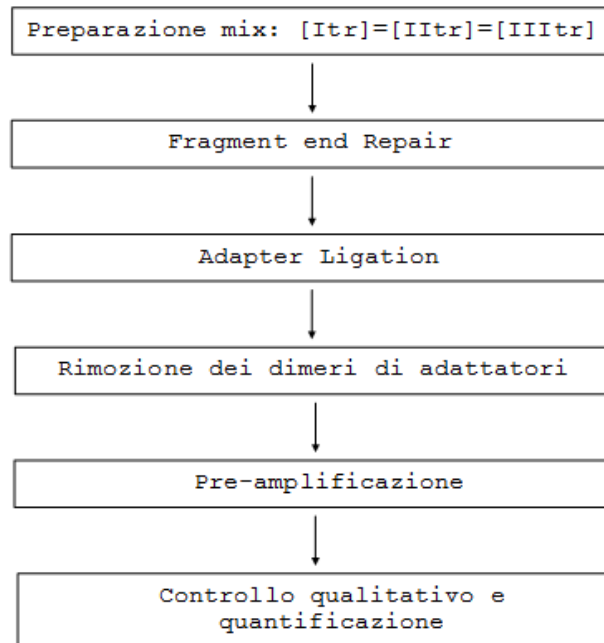
In ogni amplificazione è stato inserito anche un controllo negativo così da monitorare l'eventuale introduzione di contaminanti durante l'allestimento di ciascuna reazione.

Tutti i prodotti di PCR sono stati controllati tramite elettroforesi su gel d'agarosio al 2%. I frammenti della giusta dimensione sono stati rimossi dal gel con un bisturi, filtrati e purificati. Gli ampliconi sono stati conservati a una temperatura di -20°C sino alla preparazione delle librerie. Al contrario, i campioni che non hanno restituito bande o per i quali l'amplificazione è stata aspecifica, sono stati sottoposti ad altre reazioni di PCR e nuovamente controllati su gel.

### **3.2.4 Preparazione delle librerie**

Questa e le successive fasi sperimentali (emPCR e sequenziamento) sono state svolte presso l'Istituto di Tecnologie Biomediche del CNR, nella sede di Milano.

Per ciascun campione, la libreria è stata fatta su una mix costituita da tutti e tre i tratti in cui è stata suddivisa l'HVR1, presenti in egual concentrazione. Abbiamo stabilito una concentrazione minima di 1 ng per ciascun tratto, poiché abbiamo testato empiricamente che è la quantità soglia per avere una libreria di buona qualità. Per prima cosa abbiamo quindi quantificato ogni singolo amplificato con Agilent 2100 Bioanalyzer, High Sensitivity Kit, secondo il protocollo riportato dalla casa produttrice. Successivamente è stata preparata la mix dei tre tratti in un volume finale di 16 µl, sulla quale è stata fatta la libreria 454 seguendo il protocollo Rapid (Fig 3.2), al quale sono state apportate delle modifiche per ottimizzare la procedura nel caso del DNA antico. In particolare abbiamo utilizzato adattatori con MID diversi per ciascun campione, in modo tale da poter sequenziare più librerie contemporaneamente nella stessa lane, mantenendo l'identità di ogni campione. Rispetto a quanto indicato sul protocollo Roche, abbiamo ridotto la quantità di adattatori utilizzati: abbiamo riscontrato infatti, che la maggiore concentrazione di questi ultimi rispetto a quella del DNA endogeno favorisce la formazione di dimeri a discapito di una corretta ligazione degli adattatori alle molecole target.



**Fig. 3.2:** workflow per la preparazione delle librerie secondo il protocollo Rapid.

Prima di passare alla PCR in emulsione, ogni libreria è stata sottoposta a 20 cicli di amplificazione così da incrementare il numero di molecole a disposizione per il sequenziamento. Abbiamo utilizzato una coppia di primer complementare agli adattatori, in modo da mantenere la loro sequenza invariata durante la reazione di PCR e aumentare la specificità della reazione. Nella pre-amplificazione sono stati utilizzati 5  $\mu$ l di libreria purificata con biglie Agentcourt AMPure® XP [(Volume di biglie per reazione) = 1.2x (Volume di reazione)], per un miscela di reazione così costituita: 1X High Fidelity PCR Buffer (Invitrogen), 0.2 mM di ciascun dNTP, 2 mM MgSO<sub>4</sub>, 0.2  $\mu$ M di ciascun primer (Tabella 3.5) e 1 U di Platinum® Taq High Fidelity, per un volume finale di 50  $\mu$ l. La reazione è stata svolta con il seguente profilo:

- 94°C per 2 minuti
- ripetizione del seguente ciclo per 20 volte:
  - 94°C per 30 secondi
  - 55°C per 30 secondi
  - 68°C per 30 secondi
- 72°C per 2 minuti
- 4°C forever

Nome	Sequenza
Primer A	CCA TCT CAT CCC TGC GTG TC
Primer B	CCT ATC CCC TGT GTG CCT TG

**Tabella 3.5:** primers utilizzati per la pre-amplificazione delle librerie, complementari alle sequenze degli adattatori.

Le librerie pre-amplificate sono state purificate con biglie Agentcourt AMPure® XP e quantificate con Agilent 2100 Bioanalyzer, High Sensitivity Kit, secondo il protocollo riportato dalla casa produttrice.

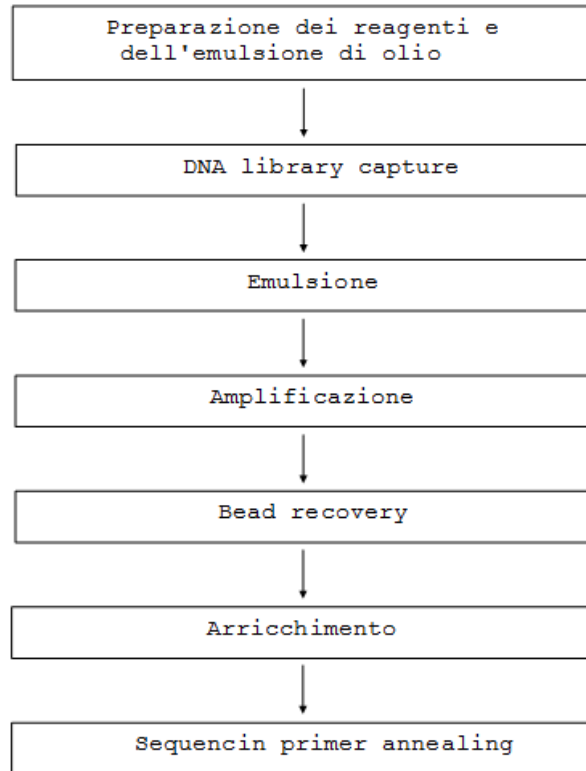
### 3.2.5 Sequenziamento delle librerie

a) PCR emulsioclonale (emPCR), recupero delle biglie e arricchimento: una volta quantificate, le librerie sono state sottoposte a PCR emulsioclonale. In una prima fase, la quantità di libreria da utilizzare è stata calcolata considerando 1 molecola di DNA per biglia. Il protocollo utilizzato è quello fornito da Roche per l'amplificazione di librerie Rapid in un elevato volume di emulsione (LV) e successivo sequenziamento su piattaforma FLX (emPCR Method – Lib-L SV), (Fig. 3.3).

Una volta arricchite, prima di passare all'annealing del primer di sequenziamento, le biglie ottenute sono state contate – con Multisizer™3

Coulter Counter®, Beckman – per valutare la resa della reazione di amplificazione: per quei campioni in cui la resa è superiore al 20% o inferiore al 3%, sono state ripetute le amplificazioni, rispettivamente riducendo o aumentando il volume di libreria utilizzata.

Le biglie arricchite sono state conservate a +4°C sino al sequenziamento.



**Fig. 3.3:** workflow dall'allestimento della emPCR all'annealing del primer di sequenziamento.

b) Sequenziamento su piattaforma 454 Genome Sequencer FLX Titanium Series (Roche): la PicoTiterPlate (PTP) è stata suddivisa in 8 lane e, per ciascun campione, sono state caricate dalle 4 alle 8 mila biglie, unendo un massimo di sei campioni per lane.

La preparazione dello strumento e della PTP è stata svolta secondo il protocollo fornito dalla Roche (Fig. 3.4).

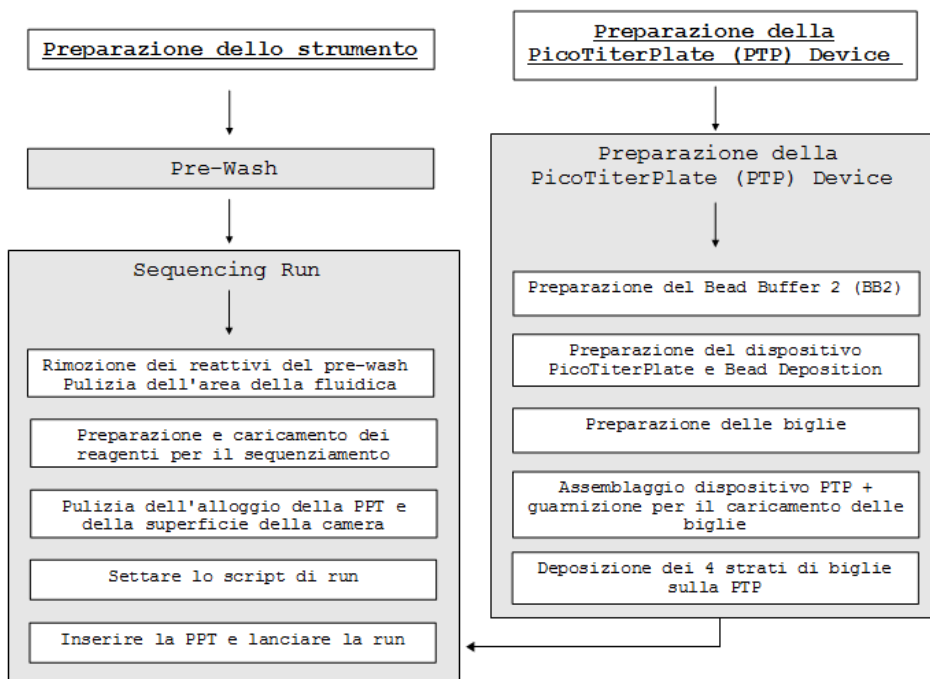


Fig 3.4: workflow per la preparazione di una run di sequenziamento su piattaforma *Genome Sequencer FLX*.

### 3.2.6 Mappaggio delle reads e ricostruzione della sequenza consenso

Le reads ottenute sono state processate con uno script progettato da noi appositamente per l'analisi degli ampliconi e per automatizzare la chiamata delle mutazioni. Ci siamo posti un limite minimo di copertura pari a 500X: i campioni che hanno raggiunto e/o superato questa soglia sono stati mappati; per gli altri è stato ripetuto il sequenziamento sino all'ottenimento di questa copertura.

Per prima cosa viene effettuata l'identificazione dei MID e il *demultiplexing* per ciascuna lane, in modo che ogni singola read venga assegnata al campione corrispondente. Il passo successivo è la rimozione dei primers (*primers trimming*) presenti alle estremità delle reads: è stata creata una lista con le sequenze dei primers utilizzati per l'amplificazione dei tre tratti, così che lo script possa ricercare quelle serie nucleotidiche (mediante "Blast") e i tagli. Per questa fase abbiamo impostato i seguenti parametri:

- *Mismatch*: non è necessaria una corrispondenza del 100% tra la sequenza dei primers inseriti nella lista e quella alle estremità delle reads, ma sono ammessi sino a 3 mismatch;

- *Match length*: la lunghezza minima del match sequenza primer/sequenza primer sulle reads, deve essere pari a 10 bp;
- “*Dual*” *search*: il primer deve essere presente ad entrambe le estremità, in questo modo ho una selezione delle reads molto più accurata.

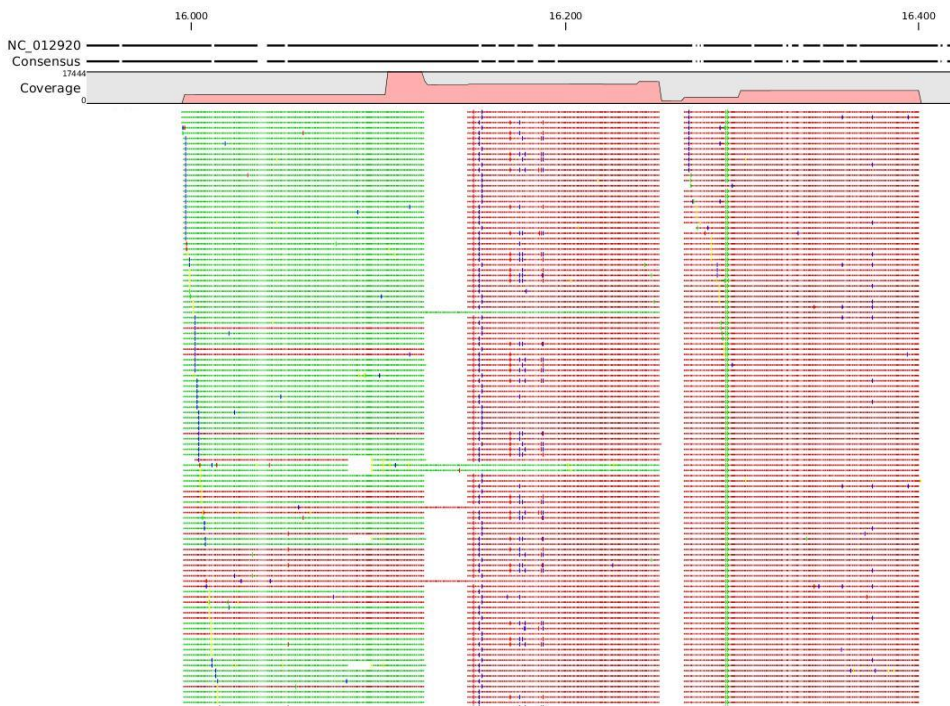
Alla fine di questo processo, per ciascun campione ho le reads filtrate pronte per il mappaggio, svolto mediante BVA. La particolarità del nostro script sta nel fatto che il mappaggio viene effettuato non sull'intera sequenza di riferimento, ma sulle tre porzioni singole in cui l'HVR1 è stata suddivisa in fase di amplificazione: in questo modo possiamo raggruppare le reads suddividendole per tratto.

Finito il mappaggio, il file in formato BAM viene processato con “Samtools”, un algoritmo che effettua la chiamata e la conta per ogni posizione nucleotidica. In questa fase abbiamo impostato i seguenti parametri (Fig. 3.5):

- viene chiamata la mutazione se più dell'80% delle *reads* portano l'allele variato;
- viene considerato un caso di eteroplasmia se la percentuale tra l'allele wild type e il variato si aggira attorno al 70-50%;
- viene identificata come *miscoding* la variante presente in meno del 15% delle reads e se si ha C→T, G→A, concentrate soprattutto alle estremità.

Nella tabella delle chiamate, lo script segnala altre informazioni importanti, come le varianti in regioni omopolimeriche (che solitamente non sono mutazioni originali ma errori di sequenziamento), le varianti in corrispondenza di inserzioni e delezioni, il rapporto tra sequenze in forward e sequenze in reverse e la frequenza allelica tra allele mutato e allele wild type (Fig. 3.5). In questo modo abbiamo ottenuto delle sequenze consenso altamente informative e di buona qualità.

Le reads filtrate possono essere mappate anche con “CLC Genomics Workbench”, su l'intera sequenza di riferimento (Fig. 3.6): abbiamo utilizzato questo software semplicemente per visualizzare l'allineamento per quei casi in cui la chiamata della variante era dubbia.



**Fig. 3.6:** mappaggio delle reads con CLC Genomics Workbench. Le reads rosse sono quelle in reverse, le reads verdi sono quelle in forward.

### 3.3 ANALISI DEI DATI

#### 3.3.1 Analisi a livello intra-popolazionistico

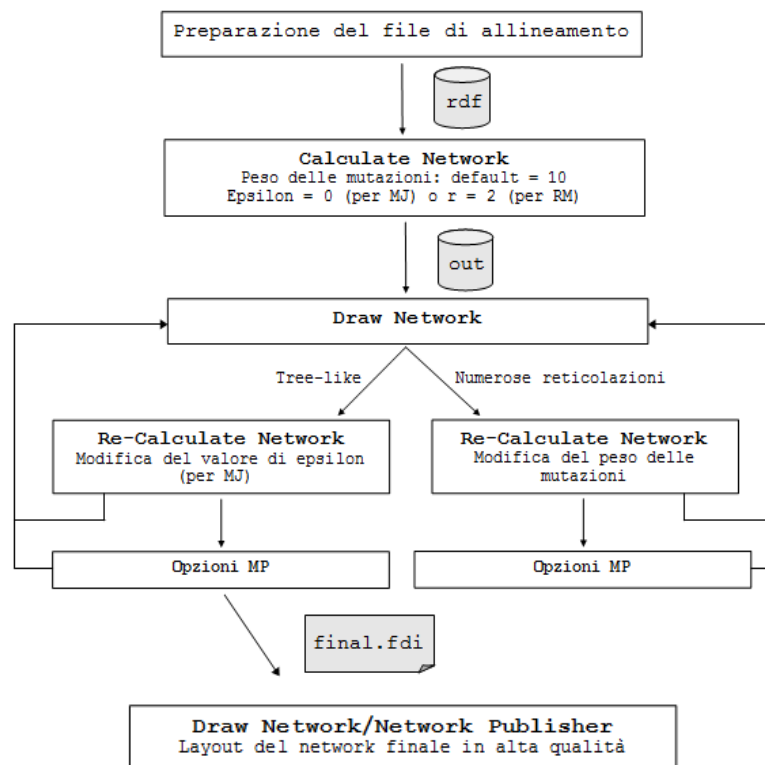
Al fine di valutare la struttura genetica dei campioni oggetto di studio, sono stati stimati alcuni parametri descrittivi della variabilità genetica attraverso l'analisi del frammento della prima regione ipervariabile dell'mtDNA dalla posizione nucleotidica 16024 alle 16384.

#### Generazione dei network

I network vengono comunemente utilizzati nelle ricostruzioni filogenetiche per identificare le varianti ancestrali, le dinamiche evolutive e per stimare i tempi d'insorgenza dei cambiamenti osservati nel campione.



Il programma scelto per la generazione dei networks è “Network 4.6.1.1”, che utilizza algoritmi specifici per i sistemi non ricombinanti, come appunto l’mtDNA. Tale software è stato sviluppato per produrre tutti i possibili alberi filogenetici di massima parsimonia da un dato data set, scegliendo tra due diverse opzioni di network-building: (i) algoritmo *reduced median* o RM network, richiede dati binari – esempio: nella posizione nucleotidica 16092 ogni taxon può avere o la T o la C. Per consentire l’interpretazione di dati complessi, è disponibile un parametro di riduzione, *r*; se la threshold di *r* è impostata su un valore sufficientemente elevato, RM produrrà un *median network* completo contenente tutti gli alberi di massima parsimonia. (ii) Algoritmo *median-joining* o MJ network, consente l’utilizzo di dati multi-state – esempio: la posizione nucleotidica 16092 può presentare indistintamente tutte e quattro le basi azotate, o un’ambiguità come la N. Quando si hanno molti dati da processare è possibile abbassare il valore di epsilon così da velocizzare il calcolo dei networks; se si aumenta il valore di epsilon, aumenta la risoluzione dei networks ma si allungano i tempi di esecuzione. Nella figura 3.7 è illustrato il workflow per il calcolo di un network RM e MJ.



**Fig. 3.7:** panoramica generale del flusso di lavoro per la generazione di network MJ e RM.

Per dati relativi alla regione di controllo dell'mtDNA, in cui si riscontrano frequenti omoplasie – mutazioni parallele o reversioni – e per cui è assente la ricombinazione, viene utilizzato generalmente il Median-Joining Network.

### Indici di diversità standard e genetica

- Numero di aplotipi: è un parametro descrittivo della diversità interna alla popolazione. È il numero dei diversi aplotipi osservati, che può essere pari o inferiore dal numero di individui.

- Diversità genetica (Nei, 1987): è detta anche eterozigosi ed è definita come la probabilità che due aplotipi scelti a caso siano differenti all'interno del campione. Viene calcolata con la seguente equazione:

$$H = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2\right)$$

dove  $n$  è il numero di sequenze nel campione,  $k$  è il numero dei diversi aplotipi presenti e  $p_i$  è la frequenza nel campione dell'aplotipo  $i$ -esimo.

La varianza è data da:

$$V(H) = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[ \sum_{i=1}^k p_i^3 - \left( \sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left( \sum_{i=1}^k p_i^2 \right)^2 \right\}$$

- Numero di siti utilizzabili: è definito come il numero di siti che sono utilizzati nel confronto a coppie tra individui della stessa popolazione e tra individui appartenenti a popolazioni diverse.

- Numero di siti polimorfici (S): è definito come il numero di loci utilizzabili che presentano più di un allele per locus.

- Numero medio di differenze a coppie ( $\pi$ ): è il numero medio di differenze tra tutte le coppie di aplotipi del campione. È dato da:

$$\pi = \frac{n}{n-1} \sum_{i=1}^k \sum_{j=1}^k p_i p_j d_{ij}$$

dove  $d_{ij}$  è la stima del numero di mutazioni verificatesi dalla divergenza tra i due aplotipi  $i$  e  $j$ ,  $k$  è il numero di aplotipi e  $p_i$  è la frequenza dell'aplotipo  $i$  e  $n$  è la dimensione del campione (Tajima 1983; Tajima 1993).

### 3.3.2 Analisi a livello inter-popolazionistico

#### Distanza genetica tra popolazioni

Per costruire le matrici di distanze genetiche è stato utilizzato il modello Kimura 2-parametri (Kimura, 1980; Jin & Nei, 1990) che fornisce una percentuale dei nucleotidi per cui due aplotipi differiscono, corretta tenendo conto della diversa probabilità di transizioni e trasversioni. Secondo tale modello la probabilità che tutte le sostituzioni abbiano luogo casualmente è, nella maggior parte dei casi, poco realistica: infatti è stato dimostrato empiricamente che le transizioni sono generalmente più frequenti delle trasversioni.

Il modello due-parametri proposto da Kimura assume che esistano delle differenze nei tassi di sostituzione tra le transizioni e le trasversioni; le transizioni hanno un tasso  $\alpha$  per sito per anno che differisce dal tasso  $\beta$  delle trasversioni. Inoltre il modello considera la possibilità di più sostituzioni per sito.

E' stata inoltre utilizzata la correzione gamma, in quanto i tassi di mutazione non possono essere assunti come uniformi per tutti i siti. Inizialmente questa correzione veniva utilizzata per i tassi di mutazione delle sostituzioni amminoacidi, ma successivamente è stata applicata anche alla regione di controllo dell' mtDNA. La forma di questa distribuzione è controllata principalmente dal parametro  $\gamma$ : più piccolo è il coefficiente  $\gamma$  e più sono assunti irregolari i tassi di mutazione. Solitamente, per l' mtDNA umano viene utilizzato un coefficiente  $\gamma$  pari a 0.26; se il tasso di mutazione fosse uguale per tutti i siti, il valore di  $\alpha$  tenderebbe all'infinito.

- Indice Fst: per misurare la distanza genetica è stato utilizzato l'indice Fst, definito come la varianza osservata di una frequenza allelica, divisa per il suo valore massimo possibile, secondo la formula:

$$F_{st} = V_p / p(1-p)$$

Può essere usato per coppie di popolazioni come misura della distanza genetica tra esse, con:

$V_p$  = varianza delle frequenze alleliche tra le due popolazioni

e

$p$  = frequenza allelica media tra le popolazioni.

Il valore ottenuto varia tra 0 e 1. Quanto più si avvicina a 1, tanto più le popolazioni risultano tra loro differenziate; al contrario, quanto più si avvicina a 0, tanto più le popolazioni saranno simili.

- Numero medio di differenze tra popolazioni: come misura della distanza genetica tra popolazioni è stato utilizzato anche il numero di differenze nucleotidiche indicate dagli indici  $D$  e  $D_A$  di Nei (Nei & Li, 1979).

$D$  rappresenta il numero grezzo, ottenuto calcolando il numero medio di differenze a coppie di sequenze tra due popolazioni, secondo l'equazione:

$$D = \pi_{12} = \sum_{i=1}^k \sum_{j=1}^{k'} x_{1i} x_{2j} \delta_{ij}$$

$D_A$  è l'indice corretto tenendo conto anche della variabilità interna di ciascuna delle due popolazioni, secondo la formula:

$$D_A = \pi_{12} - \frac{\pi_1 + \pi_2}{2}$$

dove  $k$  e  $k'$  sono il numero di aplotipi distinti nelle popolazioni 1 e 2 rispettivamente;  $x_{1i}$  è l'a frequenza dell'aplotipo  $i$ -esimo nella popolazione 1 e  $\delta_{ij}$  è il numero di differenze tra l'aplotipo  $i$  e quello  $j$ .

## Multidimensional Scaling (MDS)

Per rappresentare graficamente la matrice di distanze genetiche calcolate secondo l'indice  $F_{st}$ , è stato scelto il metodo Multidimensional Scaling. L'MDS può essere considerato un'alternativa all'analisi classica dei dati basata sulla realizzazione di alberi filogenetici. Si tratta di una tecnica esplorativa di dati che permette di ottenere una rappresentazione di "n" oggetti in uno spazio di "k" dimensioni, derivate da informazioni relative alla similarità/dissimilarità tra ciascuna coppia di oggetti. Tecnica iterativa usata per visualizzare le distanze tra oggetti in uno spazio (2D o 3D), fornisce una rappresentazione geometrica dei rapporti tra questi e permette di trovare una configurazione che approssima nel miglior modo possibile le distanze osservate. In termini più tecnici, il programma usa un algoritmo di minimizzazione di funzioni che valuta differenti configurazioni con l'obiettivo di massimizzare la bontà dell'adattamento (o minimizzare la "mancanza di adattamento") rispetto ai dati. La misura più comune usata per valutare quanto una particolare configurazione riproduce la matrice di distanze osservata è la misura di stress. Il valore di stress grezzo (o semplice)  $\Phi$  ( $\phi$ ) per una configurazione è definito da:

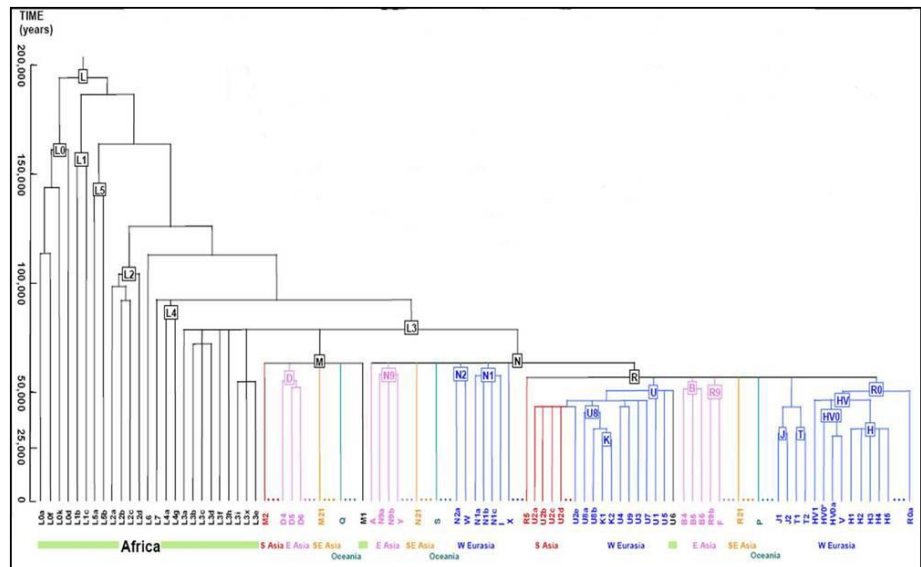
$$\Phi = \sum [ d_{ij} - f(\delta_{ij}) ]^2$$

Dove,  $d_{ij}$  rappresentano le distanze riprodotte e  $\delta_{ij}$  rappresentano invece i dati originali, cioè le distanze osservate. L'espressione  $f(\delta_{ij})$  indica una trasformazione monotona, non metrica, dei dati originali osservati (distanze). Quindi, il programma tenterà di riprodurre l'ordinamento delle distanze tra gli oggetti nell'analisi. I calcoli e la rappresentazione grafica per m'MDS sono state effettuate con il software STATISTICA 7.1 prodotto da StatSoft. Inc.

## 4. RISULTATI

Dei 103 campioni sottoposti ad analisi è stato possibile sequenziare l'intera HVR1 di 54 individui. La regione meglio rappresentata risulta essere l'Umbria, con 39 individui; per la Puglia, dei 31 campioni analizzati, solo 9 hanno restituito un profilo completo. La regione meno rappresentata risulta essere la Sardegna, con solo 6 sequenze dell'HVR1 complete: risultato questo, atteso, dato il pessimo stato di conservazione dei reperti scheletrici. Nella tabella A.1 in appendice 1, sono indicati i campioni, suddivisi per sito e le fasi sperimentali per cui hanno dato risultato. Per le successive analisi popolazionistiche abbiamo deciso di concentrarci solo sui campioni umbri, dal momento che sono quelli per cui abbiamo a disposizione il più alto numero di sequenze.

I polimorfismi di sequenza riscontrati nella regione HVR1 dei campioni analizzati ci hanno consentito di associare ciascun aplotipo alla rispettiva linea mitocondriale: il sequenziamento completo di numerosi genomi mitocondriali ha permesso, negli ultimi anni, di chiarire e definire la filogenesi di un numero sempre maggiore di aplogruppi mitocondriali che rappresentano i rami dell'albero filogenetico schematico riportato in Figura 4.1. Questi dati hanno reso più veloce e immediata la classificazione dei campioni in aplogruppi, permettendo di conoscere i motivi diagnostici sia nella regione di controllo che in quella codificante. Al momento è spesso possibile capire, mediante la sola analisi del motivo di sequenza di parte della regione di controllo, l'esatta (o probabile) appartenenza di un mtDNA non solo ad un aplogruppo, ma ad uno specifico sotto-aplogruppo. Nel nostro caso specifico, data la breve lunghezza di questa porzione di mitocondrio, nonostante l'alta risoluzione di sequenza, l'associazione con gli aplogruppi è risultata talvolta incerta; per questi casi andremo ad approfondire l'indagine con l'identificazione delle posizioni diagnostiche della porzione codificante e/o con il sequenziamento dell'intero genoma mitocondriale.



**Fig. 4.1:** albero filogenetico mondiale dei principali aplogruppi mitocondriali umani.

Nonostante le precauzioni prese per eliminare il più possibile i contaminanti, durante l'identificazione degli aplogruppi, è emerso che alcuni campioni presentavano un profilo misto derivato da più DNA mitocondriali: questi campioni, quindi, non sono stati inclusi nelle analisi popolazionistiche.

Nei successivi paragrafi vengono descritti i risultati ottenuti per ciascuna regione.

#### 4.1 SARDEGNA

Date le pessime condizioni di conservazione dei reperti ossei sardi (Fig. 4.2), tutte le fasi sperimentali sono risultate estremamente difficoltose, a partire dalla pulizia e polverizzazione sino all'amplificazione dei tre frammenti dell'HVR1.



**Fig 4.2:** campione sardo SZ114 del Su Forru de is Sinzurreddus. Si noti l'estrema frammentazione del campione e il pessimo stato di conservazione.

Dopo circa 50 prove di amplificazione in cui sono stati testati volumi diversi di estratto a diverse diluizioni per ridurre la presenza di inibitori della polimerasi, e utilizzando anche coppie di primers alternative a quelle descritte nel capitolo "MATERIALI E METODI", la regione HVR1 completa è stata ottenuta in soli 7 campioni (Tabella A.1 in appendice 1). Le successive fasi di preparazione delle librerie, emPCR e sequenziamento sono state decisamente meno problematiche, e siamo riusciti ad ottenere le sequenze di 6 campioni; per uno solo dei 7 con HVR1 completa (CAR\_H10), il sequenziamento ha sempre restituito reads aspecifiche mappanti prevalentemente su genomi batterici.

Per i campioni CAR\_H13, CAR\_H18, CAR\_H19, SI01, SZ92, SZ93, sono stati raggiunti i parametri che avevamo fissato per l'ottenimento di sequenze consenso di buona qualità, e i polimorfismi riscontrati ci hanno consentito di associare ciascun aplotipo alla rispettiva linea mitocondriale (Tabella A.2c in appendice 1).

A testimonianza dell'efficacia del nostro disegno sperimentale, è da segnalare il caso del reperto CAR\_H19: il campione presenta infatti un profilo mitocondriale frutto della miscela di più aplotipi; grazie all'elevata profondità di sequenziamento e ai parametri imposti per la chiamata delle varianti, è stato possibile identificare l'anomalia nonché il profilo dei tre mitocondri presenti (Fig. 4.3).





**Fig. 4.3:** particolare del campione CAR\_H19 con profilo derivato da una miscela di aplotipi diversi. In celeste, giallo e rosa, i tre profili mitocondriali ritrovati.

Sfortunatamente, lo scarso numero di individui che siamo riusciti a sequenziare non ci consente di effettuare analisi popolazionistiche informative, pertanto ci limitiamo a fornire solo i risultati descrittivi sopra elencati. E' da sottolineare inoltre, il fatto che anche i campioni animali appartenenti alla specie *Bos taurus* provenienti dai medesimi siti archeologici e oggetto di studio di un altro progetto, hanno avuto una resa molto bassa. Su un totale di 16 campioni sottoposti ad analisi genetiche, infatti, 4 hanno restituito un profilo mitocondriale parziale in fase di amplificazione, che una volta sequenziato ha prodotto solo reads aspecifiche; mentre per i restanti 12 reperti non è stato possibile ottenere neppure un'amplificazione. In generale, quindi, nonostante le numerose prove svolte per ottenere le informazioni genetiche necessarie al nostro obiettivo, possiamo attribuire la scarsa resa dei campioni alle caratteristiche del suolo dell'ambiente deposizionale, evidentemente inadatto ad una buona conservazione del DNA e ricco di sostanze inibitorie.

## 4.2 PUGLIA

Anche per i campioni pugliesi abbiamo avuto inaspettatamente una resa piuttosto scarsa: nonostante le buone condizioni in cui si presentavano i campioni (Fig. 4.4), dei 31 sottoposti ad analisi genetiche, solo 9 hanno restituito un profilo mitocondriale completo, mentre i restanti 22 in fase di amplificazione hanno prodotto profili parziali (Tabella A.1 in appendice 1). Anche per questa regione risulta quindi insensato svolgere analisi popolazionistiche.



**Fig. 4.4:** campione pugliese 5-T6 (Roca Vecchia)

Ad ogni modo, a ciascun campione è stato attribuito un aplotipo tranne per il campione 11-L7 che presenta un profilo misto e per il 5-T6 le cui sequenze mappano su DNA batterico (Tabella A.2b in appendice 1). 6 individui su 8 appartengono all'aplotipo H, che in effetti è il sottoinsieme più comune in Europa (40%-50%) (Richards et al., 2000; Achilli et al., 2004). L'individuo 10-T7 appartiene alla linea mitocondriale I, esclusiva dell'Europa, e presente anche nelle popolazioni pugliesi moderne (Torrioni, lavoro in preparazione). La comparsa dell'Hg I viene fissata tra i 32 e i 24 ka nella penisola balcanica da cui si è poi diffuso in tutta Europa a seguito dello scioglimento dei ghiacci dopo l'LGM. Data la particolarità di questo aplogruppo, ci proponiamo sia di ampliare il campionamento, sia di estendere l'analisi all'intero mitocondrio in modo da chiarire la filogenesi e le relazioni pugliesi antichi/pugliesi moderni.

Ci siamo anche trovati di fronte ad un caso di difficile interpretazione: il motivo che caratterizza il campione 236-T5 fa riferimento all'aplogruppo L2, che ha distribuzione esclusivamente africana (Torroni et al., 2001a). L'anomalia dovuta alla presenza di questo aplotipo in un campione pugliese antico merita ulteriori indagini mediante tipizzazione dei siti diagnostici della regione codificante o dell'intero genoma mitocondriale.

### 4.3 UMBRIA

Gli Umbri Plestini sono la popolazione meglio rappresentata, per la quale abbiamo ottenuto 39 HVR1 complete su 45 campioni a disposizione. Con il sequenziamento sono state prodotte 35 sequenze con copertura maggiore di 500X, mentre per i reperti 2, 59, 61 e 155 le reads, mappano su genomi batterici pertanto questi campioni sono stati esclusi dalle analisi.

Dati i buoni risultati, è proprio sulla regione Umbria che sono state svolte le principali analisi popolazionistiche. Come popolazioni moderne di confronto abbiamo utilizzato un database costituito da 47 popolazioni dell'Europa e del Vicino Oriente (Tabella 1.A2, in Appendice 3). Sono stati utilizzati anche i dati di sequenza della popolazione umbra moderna prodotti presso l'università di Perugia, che si è occupata del campionamento e delle fasi sperimentali sino al sequenziamento (Appendice 2).

#### 4.3.1 Analisi intra-popolazionistiche

**Variabilità mitocondriale degli Umbri Plestini.** Prima di passare alla determinazione degli aplotipi, abbiamo effettuato un ulteriore "controllo di qualità" delle sequenze, così da escludere eventuali prodotti di contaminazione. A tal proposito, le sequenze ottenute sono state confrontate con i profili degli operatori che sono venuti a contatto con i reperti: si evidenzia una condivisione del polimorfismo in posizione 16311 tra i campioni 5, 44, 79, 147, 151, 192, 198, 209 e 294 e chi ha condotto le analisi di laboratorio. Nonostante ciò, abbiamo scelto di non scartare questi campioni, per quattro motivi: non è stata riscontrata contaminazione originata nelle fasi più critiche della manipolazione dei reperti (prelievo in corso di scavo, eventuale lavaggio, studio morfologico), pertanto risulta ancora meno probabile la possibilità che la contaminazione sia avvenuta nel corso delle analisi molecolari in virtù delle numerose e costanti precauzioni adottate in laboratorio; emerge inoltre che i campioni condividono

con gli operatori esclusivamente la posizione 16311, e nei casi in cui l'operatore presenti altri siti polimorfici, questi non vengono ritrovati nei campioni "incriminati"; infine, il profilo mitocondriale in questione risulta piuttosto frequente, di conseguenza la condivisione dell'aplotipo rappresenta un evento tutt'altro che raro; eliminando dall'analisi questi campioni si sarebbe persa un'importante informazione non solo in termini di quantità (numero di campioni) ma anche in termini di qualità, vista la presenza dell'aplotipo nei campioni moderni utilizzati nel confronto.

La combinazione di mutazioni presenti nella sequenza dell'HVR1 ci ha permesso di affiliare ciascun mtDNA in esame ad uno specifico aplogruppo (Tabella A.2a in appendice 1); anche per la popolazione umbra sono stati riscontrati casi in cui il profilo deriva da un mix di aplotipi e uno in cui l'affiliazione non è certa: si tratta dell'individuo 39, che mostra il polimorfismo 16223 tipico del W, un aplogruppo piuttosto antico (datato a circa 24.000 anni fa) e presente anche nelle popolazioni umbre moderne. L'incertezza dell'affiliazione deriva dal fatto che la maggior parte delle mutazioni diagnostiche di questo aplogruppo si trovano nella regione codificante pertanto, in attesa della sequenza mitocondriale completa, il campione è stato escluso dalle analisi filogenetiche.

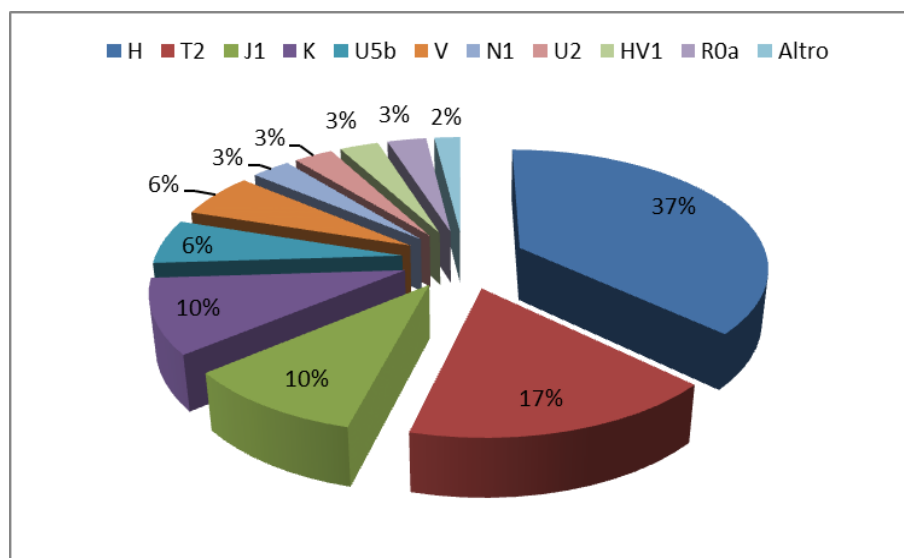
**Indici di diversità standard e di diversità genetica.** In tabella 4.1 sono riportati i valori degli indici di diversità standard e di diversità genetica più comuni, calcolati a partire dal gruppo delle 30 sequenze di Umbri Plestini ottenute.

<b>Parametri considerati</b>	<b>Valore osservato</b>
Numero di loci utilizzabili	360
Numero di siti polimorfici	36
Numero di transizioni osservate	32
Numero di trasversioni osservate	7
Numero di aplotipi	26
Diversità genetica	$0.991 \pm 0.012$
Numero medio di differenze a coppie	$4.9929 \pm 2.47$

**Tabella 4.1:** indici di diversità standard e diversità genetica dei campioni umbri antichi.

Le 30 sequenze dell'HVR1 considerate nelle analisi filogenetiche presentano 26 aplotipi diversi appartenenti a 12 linee mitocondriali. Il calcolo

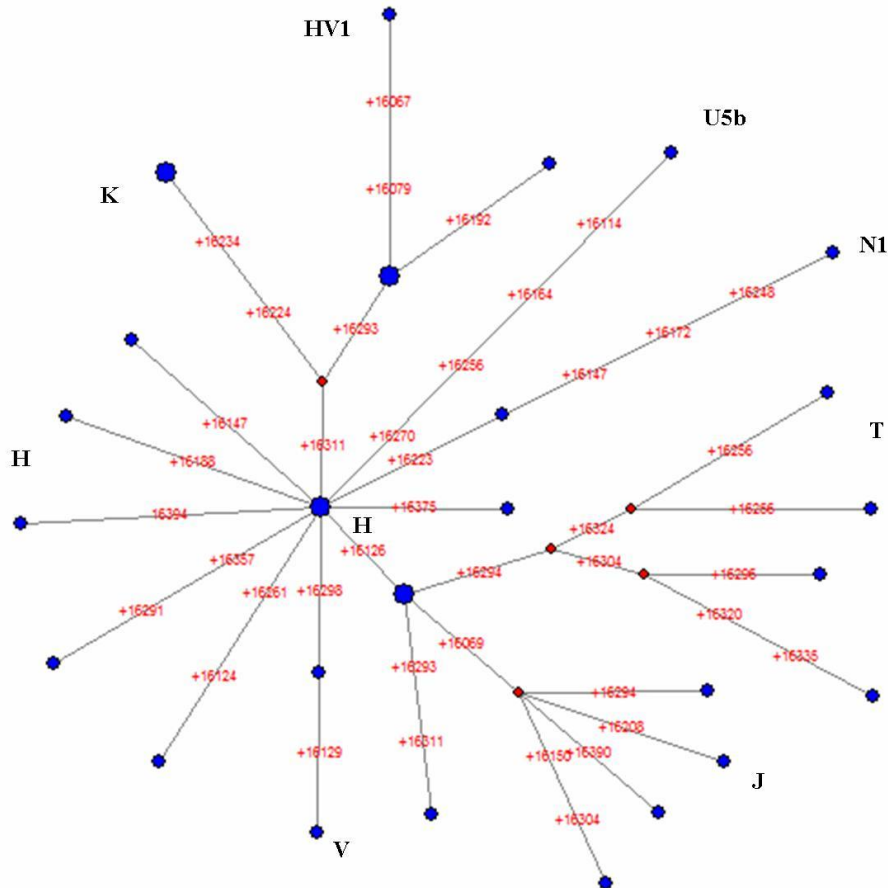
dell'indice Nei (1987) mostra per la popolazione umbra antica una diversità genetica elevata,  $Hd = 0.991 \pm 0.012$ . L'aplogruppo maggiormente rappresentato risulta essere l'H con una frequenza del 37% (Fig. 4.5), in accordo con le alte frequenze riscontrate nelle popolazioni moderne sia umbre che della parte più occidentale dell'Europa (Achilli et al., 2004). Il restante gruppo delle linee mitocondriali identificate sono quelle tipiche dell'Europa e dell'Eurasia Occidentale, come risulta evidente confrontando le Figure 4.1 e 4.5.



**Fig. 4.5:** frequenze dei principali aplogruppi mitocondriali identificati negli Umbri antichi.

È stato realizzato un *Median-Joining Network* (Fig. 4.6) per rappresentare la distribuzione delle sequenze antiche umbre e visualizzare come sono filogeneticamente collegate tra loro. Per la realizzazione del network è stata presa in considerazione la diversa frequenza di transizioni e trasversioni, così come le singole posizioni nucleotidiche sono state suddivise nelle categorie “rapido”, “intermedio” e “lento” relativamente al proprio tasso di transizione in modo da assegnare un peso inversamente proporzionale alla probabilità con cui ciascun sito nucleotidico può andare incontro a mutazione. Tale peso rispecchia l'importanza relativa che i caratteri assumono nella costruzione del network. In più sono state rimosse le mutazioni 16183 e 16189, dal momento che si trovano all'interno di una zona omopolimerica – poli-C stretch – risultando verosimilmente il prodotto di un errore di sequenziamento.

I nodi rappresentano gli aplotipi riscontrati e la loro dimensione è proporzionale alla frequenza dell'aplotipo nel campione. I numeri in rosso indicano le posizioni nucleotidiche per le quali, in seguito a mutazione, si distinguono gli aplotipi L'aplotipo denominato H\_2 al centro dell'albero rappresenta la Sequenza di Riferimento di Cambridge (CRS) e dalla topologia dell'albero emerge una distribuzione piuttosto uniforme degli aplogruppi.

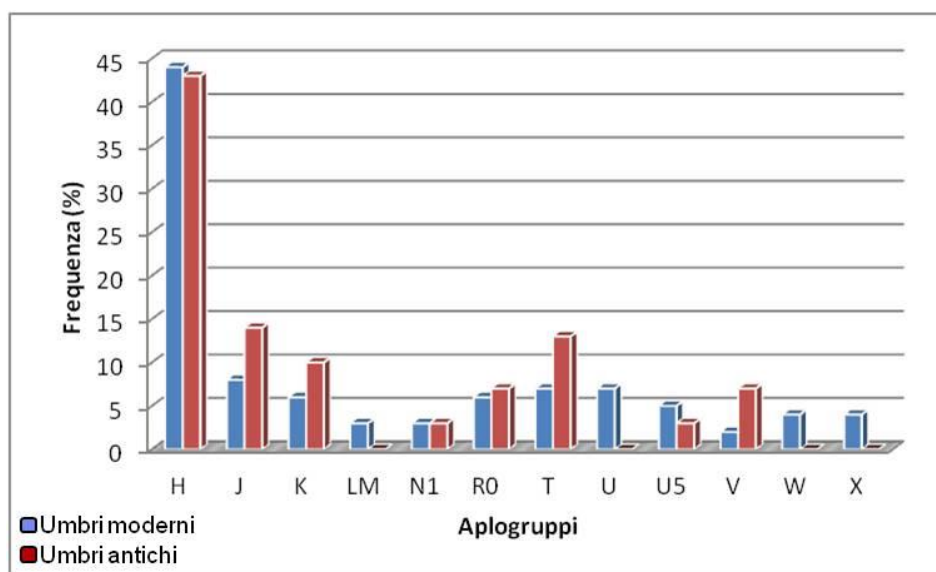


**Fig. 4.6:** Median-Joining Network delle sequenze umbre antiche

### 4.3.2 Analisi inter-popolazionistiche

E' stato realizzato un confronto tra gli Umbri Plestini e altre quattro popolazioni italiane antiche dell'Età dei Metalli: Etruschi, Messapi, Sardi (Nuragici) e Piemontesi (protostorici – Età del Bronzo) (Tabella A.4 in appendice 3), per un totale di 69 sequenze.

Inoltre, per cercare di capire quanto la popolazione antica abbia contribuito all'attuale struttura genetica dell'Umbria (Fig. 4.7), e come si colloca nel contesto di altri gruppi italiani ed europei, abbiamo confrontato il data set antico con 47 popolazioni dell'Eurasia Occidentale (Tabella A.3, in Appendice 3). Per le sequenze moderne abbiamo considerato soltanto i polimorfismi dell'HVR-1, per un frammento che va dalla posizione 16024 alla 16383 in modo da uniformare il dato moderno con quello antico.



**Fig. 4.7:** confronto “super-aplogruppi” Umbri antichi/moderni. Per gli Umbri moderni è stato utilizzato il dato di sequenza relativo solo all'HVR-1.

Già dalla Fig. 4.7 emerge come gli Umbri Plestini condividano con gli Umbri attuali solo i super-aplogruppi maggiormente rappresentati in Eurasia, mentre mancano completamente gli aplogruppi specifici, come U, X e W, presenti nella popolazione moderna con frequenze attorno al 6%. La mancanza

di continuità genetica è confermata dagli indici di Fst e dall'indice D di Nei, riportati in seguito

Gli indici di Fst sono stati calcolati tra gli Umbri Plestini e le popolazioni antiche (Tabella 4.4) e moderne (Tabella 4.5).

	Umbri Plestini	Etruschi	Messapi	Sardi	Piemontesi
Umbri Plestini	0,000				
Etruschi	<b>0,037</b>	0,000			
Messapi	0,000	<b>0,062</b>	0,000		
Sardi	<b>0,055</b>	<b>0,039</b>	<b>0,096</b>	0,000	
Piemontesi	0,003	0,030	0,075	<b>0,280</b>	0,000

**Tabella 4.4:** indici di distanza genetica Fst tra popolazioni antiche. In grassetto sono indicati gli indici con P-value significativo.

	Umbri	Aust	Basq	Belg	Brit	Bulg	Cata
Umbri Plestini	0.0356	0.0385	0.0474	0.0413	0.0385	0.0361	0.0398

	Chec	Corn	Dane	Druz	Esto	Finl	Fren
Umbri Plestini	0.0384	0.0421	0.0426	0.0338	0.0390	0.0414	0.0422

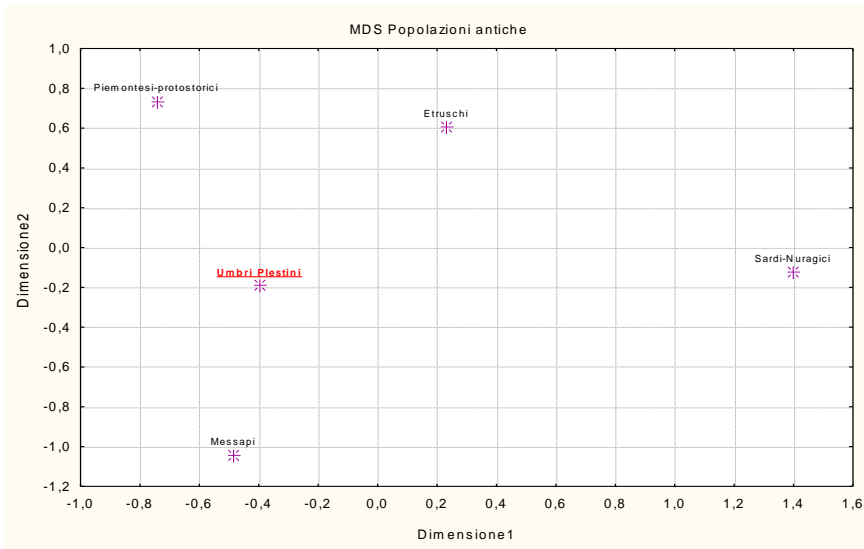
	Gale	Geor	Gern	Gers	Ingu	Kare	Kurd
Umbri Plestini	0.0467	0.0371	0.0415	0.0420	0.0388	0.0426	0.0363



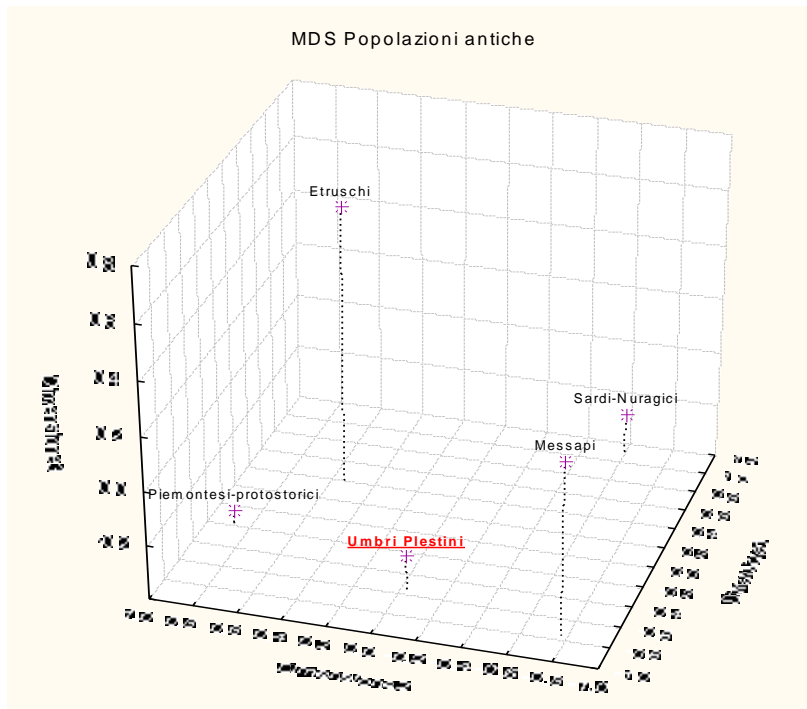
	MEast	Norw	Neth	Port	Russ	Spap	Swed
Umbri Plestini	0.0311	0.0392	0.0365	0.0423	0.0393	0.0391	0.0402
	Swid	Siri	Volg	Wale	Bologna	Campania	Lazio
Umbri Plestini	0.0423	0.0346	0.0402	0.0436	0.0383	0.0388	0.0406
	Albania	Molise	Abruzzo	Puglia	Grecia	Piem_Torino	Casentino
Umbri Plestini	0.0400	0.0406	0.0365	0.0381	0.0381	0.0361	0.0429
	Murlo	Volterra	Tuscan	Turchi	Sard ogliastra	Sard gallura	
Umbri Plestini	0.0377	0.0376	0.0354	0.0347	0.0528	0.0395	

**Tabella 4.5:** indici di distanza genetica  $F_{st}$  tra Umbri Plestini e popolazioni moderne italiane e dell'Eurasia. Tutti gli indici di P-value hanno valore significativo.

Dalle matrici di  $F_{st}$  sono stati ricavati i grafici MDS che rappresentano le distanze genetiche tra gli Umbri Plestini e le popolazioni antiche e moderne considerate (Fig 4.8; 4.9 e 4.10). Nel grafico di Fig. 4.10 (Umbri Plestini-popolazioni eurasiatiche moderne) il consistente numero di popolazioni rappresentate implica un certo grado di distorsione nella raffigurazione delle distanze genetiche, che però si riduce nella rappresentazione a tre dimensioni.



(stress = 0,005)



(stress = 0,002)

Fig 4.8: MDS dei valori di Fst tra Umbri Plestini e popolazioni italiane antiche

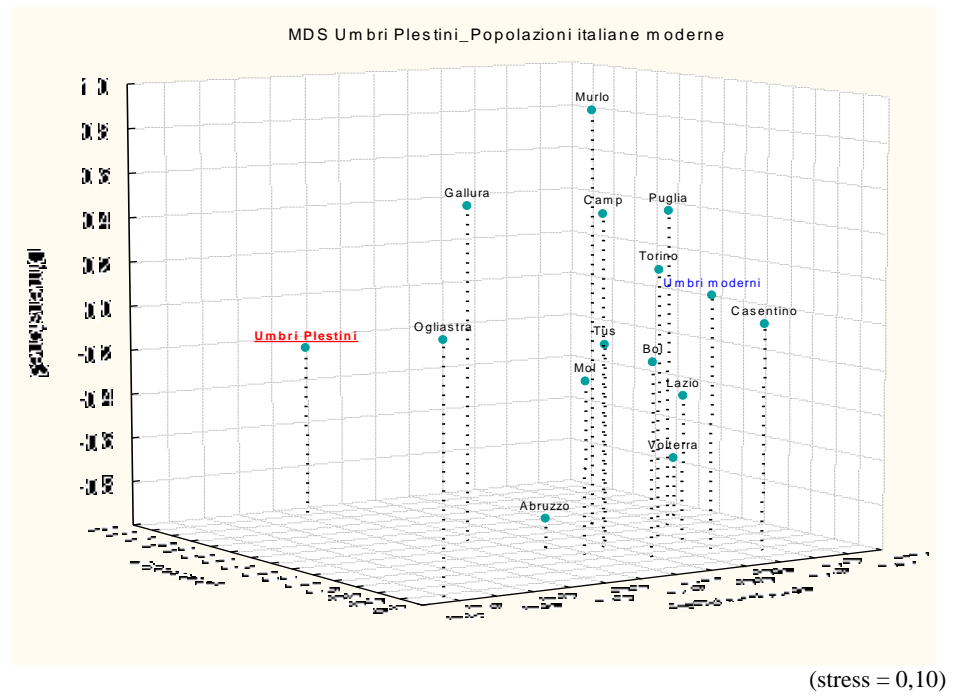
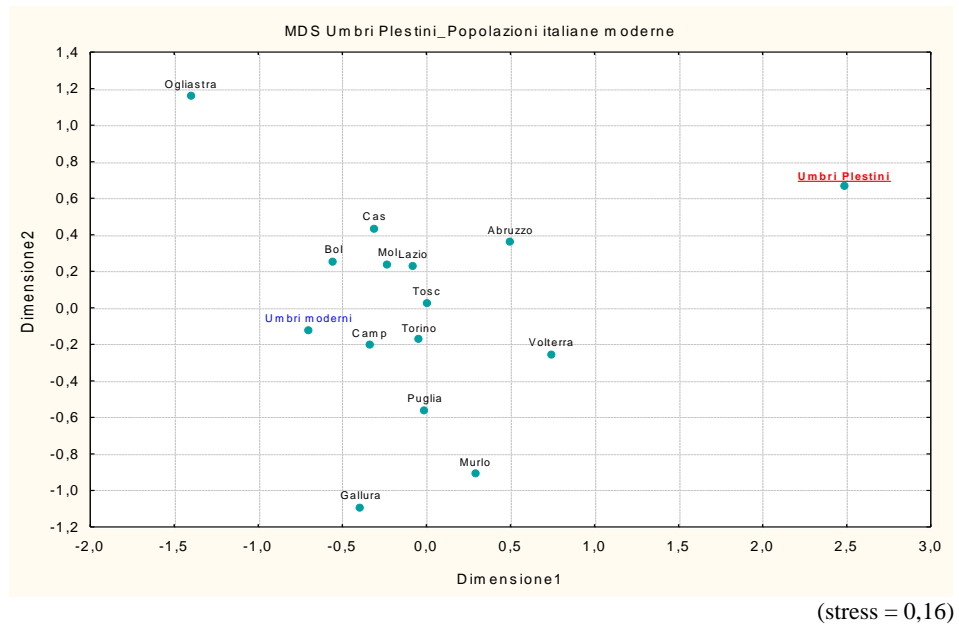
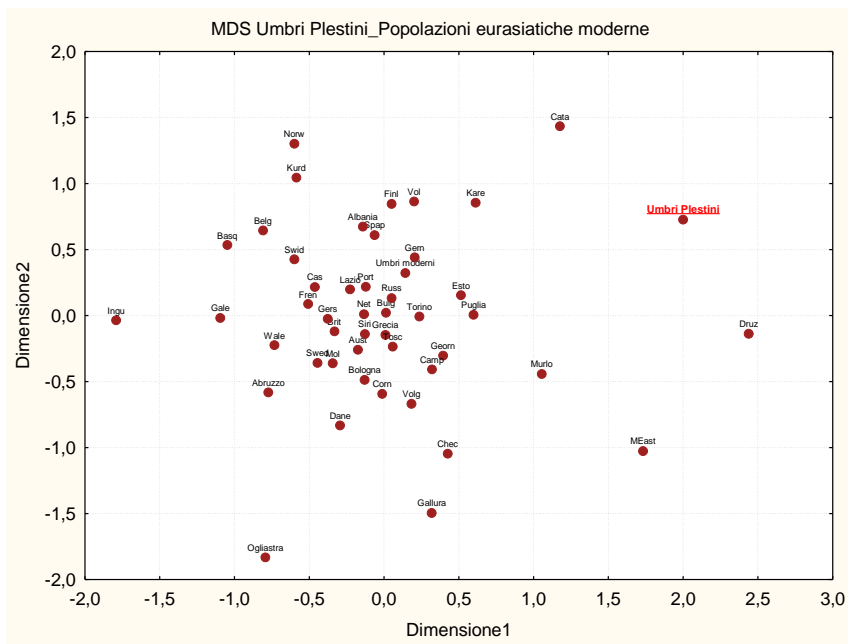
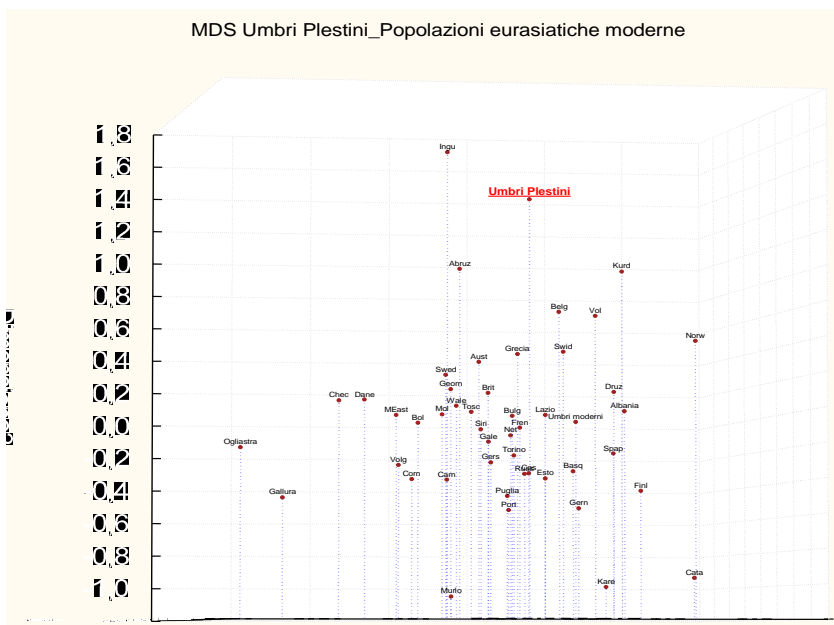


Fig 4.9: MDS dei valori di Fst tra Umbri Plestini e popolazioni italiane moderne



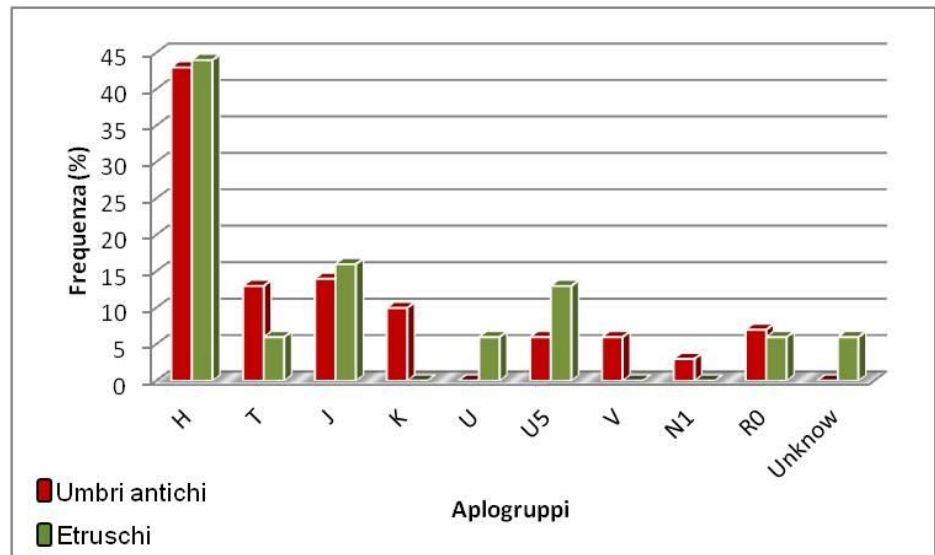
(stress = 0,22)



(stress = 0,15)

Fig. 4.10: MDS dei valori di Fst tra Umbri Pleistini e popolazioni eurasiatiche moderne

E' possibile notare come, all'interno delle popolazioni antiche considerate (Fig. 4.8) non esista una vicinanza genetica maggiore tra Umbri Plestini ed Etruschi, come invece si poteva immaginare, data l'esistenza di intensi scambi commerciali tra queste due culture attraverso la Via del Ferro; come si osserva dal plot, infatti, la distanza tra Umbri antichi ed Etruschi è equiparabile a quella che c'è tra i primi e Messapi e Piemontesi. Il dato è confermato anche dal grafico delle frequenze aplotipiche (Fig.4.11) e dal network (Fig.4.12) dai quali emerge una condivisione della variabilità tra le due popolazioni antiche, inferiore al 7%.



**Fig 4.11:** confronto “super-aplogruppi” Umbri antichi/Etruschi.



	Gale	Georn	Gern	Gers	Ingu	Kare	Kurd
Umbri Plestini	7,401	7,983	7,615	7,781	7,995	7,750	7,424
	MEast	Norw	Neth	Port	Russ	Spap	Swed
Umbri Plestini	9,157	7,066	8,236	7,472	7,791	7,712	8,176
	Swid	Siri	Volg	Wale	Bologna	Campania	Lazio
Umbri Plestini	7,575	8,730	7,956	7,417	7,732	7,928	7,460
	Albania	Molise	Abruzzo	Puglia	Grecia	Piem_Torino	Casentino
Umbri Plestini	7,969	7,599	8,042	8,063	7,749	8,082	7,371
	Murlo	Volterra	Tuscan	Turchi	Sard Ogliastra	Sard Gallura	
Umbri Plestini	8,573	8,527	8,081	7,936	7,302	8,324	

**Tabella 4.6:** numero medio di differenze a coppie tra Umbri Plestini e popolazioni moderne.

	Umbri Plestini	Etruschi	Messapi	Sardi	Piemontesi
Umbri Plestini	4,92874	4,43889	4,64074	3,37826	4,62222
Etruschi	0,16303	3,62299	4,20000	2,62609	3,86667
Messapi	0,00971	0,22184	4,33333	2,9934	4,55556
Sardi	0,21824	0,11894	0,12802	1,39130	2,92754
Piemontesi	0,15785	0,05517	0,38889	0,23188	4,00000

**Tabella 4.7:** numero medio di differenze a coppie tra Umbri Plestini e popolazioni antiche.

*Above diagonal: average number of pairwise differences between populations*

*Diagonal elements: average number of pairwise differences within populations*

*Below diagonal: corrected average pairwise differences.*

Per eliminare dal calcolo del numero medio di differenze a coppie tra popolazioni l'influenza della diversità interna di ciascuna delle popolazioni confrontate, è stato usato l'indice  $D_A$  di Nei che sottrae alla distanza la media delle variabilità interne. Nella tabella 4.8 sono indicati i valori di  $D_A$  tra Umbri Plestini e popolazioni moderne, mentre nella tabella 4.7, nella parte sottostante la diagonale, i valori di  $D_A$  tra Umbri Plestini e altre popolazioni antiche.

	Umbri	Aust	Basq	Belg	Brit	Bulg	Cata
Umbri Plestini	0,287	0,297	0,318	0,331	0,293	0,294	0,313

	Chec	Corn	Dane	Druz	Esto	Finl	Fren
Umbri Plestini	0,308	0,309	0,320	0,260	0,306	0,291	0,301

	Gale	Georn	Gern	Gers	Ingu	Kare	Kurd
Umbri Plestini	0,317	0,290	0,298	0,305	0,306	0,314	0,268

	MEast	Norw	Neth	Port	Russ	Spap	Swed
Umbri Plestini	0,270	0,275	0,296	0,303	0,282	0,289	0,326

	Swid	Siri	Volg	Wale	Bologna	Campania	Lazio
Umbri Plestini	0,304	0,304	0,316	0,300	0,285	0,302	0,292

	Albania	Molise	Abruzzo	Puglia	Grecia	Piem_Torino	Casentino
Umbri Plestini	0,313	0,297	0,296	0,306	0,289	0,288	0,288



	Murlo	Volterra	Tuscan	Turchi	Sard Ogliastra	Sard Gallura
Umbri Plestini	0,322	0,320	0,284	0,274	0,333	0,327

**Tabella 4.8:** numero di differenze a coppie tra Umbri Plestini e popolazioni eurasiatiche moderne indicato dall'indice  $D_A$  di Nei.

I valori di  $D_A$  sono più o meno uniformi tra gli Umbri Plestini e le popolazioni antiche e tra i primi e le popolazioni eurasiatiche moderne (comprese quelle italiane), trattandosi peraltro di valori piuttosto alti, indici di un'elevata distanza genetica: ancora una volta quindi, emerge la peculiarità della popolazione plestina, che pare discostarsi sia dai gruppi del passato che da quelli attuali. Emerge inoltre, come la vicinanza dei Plestini con gli Umbri moderni ( $D_A = 0,287$ ) non sia maggiore di quella che i primi hanno con le altre popolazioni italiane ed europee in generale.

## 5. DISCUSSIONE E CONCLUSIONI

Questo lavoro ha avuto lo scopo di inquadrare le dinamiche popolazionistiche della nostra penisola durante l'Età dei Metalli – epoca di particolare interesse, dato il fiorire di numerose culture locali – e individuare se e in che misura le popolazioni dell'epoca abbiano contribuito alla struttura genetica dell'Italia attuale. Sfortunatamente, delle 3 regioni di partenza solo l'Umbria ha fornito il più alto numero di sequenze, pertanto le indagini popolazionistiche si sono concentrate esclusivamente sugli Umbri antichi.

### 5.1 CONSIDERAZIONI FILOGENETICHE

Ci siamo proposti quindi, di studiare alcune questioni riguardanti la storia degli Umbri Plestini, una popolazione mai stata studiata prima d'ora, ma sicuramente interessante data la sua collocazione geografica e gli intensi scambi commerciali sia con le popolazioni confinanti che con quelle più lontane (Etruschi, Piceni, Greci e in seguito anche Romani), come testimoniano i corredi funerari delle necropoli. I punti a cui abbiamo cercato di dare una risposta sono: (i) caratteristiche genetiche degli Umbri Plestini; (ii) continuità genetica tra Umbri Plestini e moderni; (iii) relazione Plestini-Etruschi; (iv) origine della popolazione plestina.

Il gruppo di campioni analizzati, rappresentato da 30 individui, mostra un indice di diversità genetica piuttosto elevato ( $Hd = 0.991 \pm 0.012$ ) e un numero medio di differenze a coppie paragonabile alla media delle popolazioni moderne piuttosto che a quella delle popolazioni antiche sinora studiate. Questi dati suggeriscono l'esistenza di una popolazione numericamente ampia, perciò geneticamente variabile, confermata anche dalle testimonianze archeologiche: agli inizi del VI secolo a.C. infatti si riscontra un significativo aumento demografico con un'occupazione del territorio di 120-130 Km<sup>2</sup>.

Dai nostri dati emerge inoltre una certa peculiarità genetica degli Umbri Plestini: dagli indici di  $F_{st}$  (Tabella 4.5) è abbastanza chiaro come la popolazione dei Plestini si discosti sensibilmente dalle popolazioni moderne dell'Italia centrale, compresi gli Umbri moderni. Secondo questi risultati è da

escludere pertanto un consistente contributo plestino al pool genico delle popolazioni moderne. Il dato è confermato anche dalle analisi filogeografiche condotte sulla popolazione umbra attuale: dalla figura 4.A2 (Allegato 2) risulta evidente la vicinanza tra gli Umbri e le popolazioni della penisola Iberica, al contrario dei Plestini che invece si discostano da tutte le popolazioni dell'Europa occidentale (MDS di Fig. 4.10).

Allo stesso modo, una volta inseriti nel contesto eurasiatico (Fig. 4.10 e Tabella 5.5), gli Umbri Plestini mantengono ancora una loro identità genetica non mostrando alcuna vicinanza con nessuna delle popolazioni europee e del Vicino Oriente, come indica anche il numero di differenze a coppie di sequenze ( $D_A$ ) il cui valore è pressoché uniforme per tutte le popolazioni considerate. Da questi risultati, peraltro piuttosto inattesi, risulta difficile individuare un centro d'origine della popolazione plestina, almeno tra le aree geografiche considerate. Data la sua lontananza da tutte le popolazioni eurasiatiche, si può ipotizzare un'origine autoctona da gruppi più antichi (paleolitici? Primi coloni neolitici?): considerato il relativo basso numero di individui tipizzati e la corta regione genetica analizzata, risulta difficile supportare questa visione. Sicuramente l'incremento del numero di individui studiati, il sequenziamento dell'intero genoma mitocondriale e le simulazioni basate sulla coalescenza e sulla statistica bayesiana ci potranno aiutare a definire le origini di questa particolare popolazione antica.

Ancora una volta, la singolarità dei Plestini emerge quando si vanno a confrontare con le popolazioni coeve. In particolare ci siamo concentrati sulle relazioni con i vicini Etruschi, dal momento che le culture materiali testimoniano intensi scambi tra i due gruppi e una forte influenza degli Etruschi sui manufatti plestini. Come si osserva dal plot di Fig. 4.8, la distanza tra Umbri antichi ed Etruschi è equiparabile a quella che c'è tra i primi e le altre popolazioni antiche considerate e, come emerge dal network di Fig. 4.12, le due popolazioni condividono solo il 7% della variabilità riscontrata. Questi risultati quindi ci portano a ipotizzare che gli scambi tra Plestini ed Etruschi fossero solo di tipo economico e culturale, e che entrambe le popolazioni mantenessero una loro identità genetica. Ancora una volta però per confermare questa ipotesi si rende necessario il sequenziamento dell'intero mtDNA e analisi statistiche più approfondite.

Nonostante rimangano ancora dei punti da approfondire, partendo da questa prima descrizione della variabilità mitocondriale dell'Umbria antica ottenuta grazie alle analisi svolte in questa tesi, sarà possibile ipotizzare degli scenari che descrivano possibili rapporti di discendenza e dinamiche demografiche. Sarebbe pretenzioso pensare di poter fornire un quadro definitivo sulla storia demografica dell'Umbria, ma quantomeno potremo essere in grado di scartare una serie di processi demografici che si rivelino incompatibili con i dati, e quindi di restringere il novero dei processi possibili. In questo senso, le

future analisi potranno rispondere a domande del tipo: dove si colloca l'origine degli Umbri Plestini? Ci sono regioni dell'Umbria attuale che conservano tracce del patrimonio genetico plestino? Oppure essi si sono estinti senza lasciare traccia nel pool attuale? Con queste ulteriori analisi sarà possibile individuare quale modello popolazionistico, o insieme di modelli, descriva al meglio i dati osservati e riassume quindi gli eventi che hanno interessato la regione.

## 5.2 CONSIDERAZIONI METODOLOGICHE

Il nostro lavoro si è concentrato anche su alcuni aspetti metodologici mirati ad adattare i protocolli NGS (454 *Genome Sequencer* – Roche) allo studio dell'aDNA e migliorare la qualità del dato finale.

Utilizzando un approccio combinato metodologia classica/Next Generation Sequencing abbiamo potuto sfruttare i vantaggi della PCR e quelli del pirosequenziamento per ottenere un'alta ridondanza di sequenza e un'alta qualità di lettura per ciascuna posizione nucleotidica: in questo modo è stato possibile identificare inequivocabilmente le mutazioni, individuare casi di eteroplasmia e valutare la percentuale di misincorporazioni, ottenendo così delle sequenze consenso altamente informative. Questo è stato possibile grazie ad alcune modifiche apportate ai protocolli forniti dalla Roche (vedi capitolo "MATERIALI E METODI", paragrafo 3.2 e successivi) che si sono rivelate efficaci e pertanto potranno essere utilizzate anche in seguito per studi di questo tipo.

Inoltre la produzione di uno script specifico per il mappaggio delle reads da ampliconi (vedi capitolo "MATERIALI E METODI", paragrafo 3.2.6) si è rivelato fondamentale per l'interpretazione dei dati di sequenziamento e per identificare la presenza di più profili mitocondriali nei campioni analizzati (Fig. 4.3). In questo senso, lo script può risultare di utile applicazione anche in campo forense, là dove si rende spesso necessario dover discriminare tracce biologiche miste, come nei casi di stupro.

Possiamo infine concludere evidenziando come questo studio è risultato utile non solo per cercare di far luce su alcune questioni filogeografiche e filogenetiche riguardanti la storia del popolamento italiano, ma anche perché ci ha dato la possibilità di migliorare le metodologie d'indagine a disposizione per lo studio del DNA antico.

## 6. BIBLIOGRAFIA

**Achilli, 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.

**Ammerman, A.J.**, Cavalli-Sforza, L.L. (1973) A population model for the diffusion of early farming in Europe. In *The Explanation of Culture Change*, C. Renfrew, ed. (London: Duckworth).

**Ammerman, A.J.**, Cavalli-Sforza, L.L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ Press, Princeton).

**Anikovich, M.V.** et al. (2007) Early Upper Paleolithic in Eastern Europe and implications for the dispersal of modern humans. *Science* 315, 223–226.

**Avise, J.** et al. (1987) Intraspecific phylogeography: the molecular bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18, 489–522.

**Barbujani, G.**, Sokal, R.R. (1991) Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. *Am. J. Hum. Genet.* 48, 398–411.

**Barbujani, G.** et al. (1995) Geographical structuring in the mtDNA of Italians. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9171–9175.

**Bandelt, H.J.** et al., (1995) Mitochondrial portraits of human populations using median networks. *Genetics*. 141, 743-753.

**Bandelt, H.J.** et al. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37-48.

**Barker, G.** (2006) *The Agricultural Revolution in Prehistory*. (Oxford: Oxford University Press).

**Bar-Yosef, O.** (1998) The Natufian culture in the Levant, threshold to the origins of agriculture. *Evol. Anthropol* 6, 159–177.

- Benazzi, S.** et al. (2011) Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479, 525–528.
- Boattini, A.** et al. (2013) Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS One*. 8(5):e65441.
- Bollongino, R.** et al. (2012) Modern taurine cattle descended from small number of Near-Eastern founders. *Mol. Biol. Evol.* 29, 2101-2104.
- Bomoni Ponzi, L.** (1985) Topographie Survey of the Colfiorito di Foligno Plateau a Contribution toward the study of the population in the territory of the Plestini. in *BAR 1985, Papers in Italian Archeology*, 5, 201-238.
- Bos, K.I.** et al. (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506-10.
- Bradley, G.** (2000) *Ancient Umbria. State Culture and Identity in Central Italy From the Iron Age to the Augustan Era.* Published in the United States by Oxford University Press Inc., New York.
- Bramanti, B.** et al. (2009) Genetic discontinuity between local hunter-gatherers and Central Europe's first farmers. *Science* 326, 137–140.
- Brandt, G.** et al. (2013) Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. *Science* 342, 257-261.
- Briggs, A.W.** et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A* 104, 14616-14621.
- Briggs, A.W.** et al. (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325, 318-321.
- Brisighelli, F.** et al. (2012) Uniparental Markers of Contemporary Italian Population Reveals Details on Its Pre-Roman Heritage. *PLoS ONE* 7: e50794.
- Brotherton, P.** et al. (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* 35, 5717–5728.
- Brotherton, P.** et al. (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* 4:1764.

- Brown, T.A.** et al. (2008). The complex origins of domesticated crops in the Fertile Crescent. *Trends Ecol. Evol.* 24, 103–109.
- Brown, W.M.** (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3605-3609.
- Burbano, H.A.** et al. (2010) Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328, 723-725.
- Burger, J.,** Thomas, M.G. (2011) The palaeopopulationgenetics of humans, cattle and dairying in Neolithic Europe. In *Human Bioarchaeology of the Transition to Agriculture* (Pinhasi, R. and Stock, J.T., eds), pp. 371–384, Wiley & Sons.
- Campos, P.F.** et al. (2010) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc. Natl. Acad. Sci. U.S.A* 107, 5675–5680.
- Cann, R.L.** et al. (1987) Mitochondrial DNA and human evolution. *Nature* 325, 31-36.
- Capelli, C.** et al. (2007) Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Mol. Phylogenet. Evol.* 44, 228–239.
- Caramelli, D.** et al. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6593–6597.
- Caramelli, D.** et al. (2006) A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr Biol.* 16, 630-632.
- Caramelli, D.** et al. (2007) Genetic variation in prehistoric Sardinia. *Hum. Genet.* 122, 327-36.
- Caramelli, D.** et al. (2008) A 28,000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences. *PLoS One.* 3(7):e2700.
- Cauvin, J.** (2000) *The Birth of the Gods and the Origins of Agriculture* (Cambridge: Cambridge University Press).

**Cavalli-Sforza, L.L.**, Piazza, A. (1993) Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur. J. Hum. Genet.* 1, 3–18.

**Cavalli-Sforza, L.L.** et al. (1993) Demic expansions and human evolution. *Science.* 259, 639-646.

**Cavalli-Sforza, L.L.** et al. (1994) *The History and Geography of Human Genes.* Princeton University Press, Princeton, New Jersey.

**Cavallo, T.** (1999) L'acropoli di Scamnum. In "Archeo", n. 1 (167), pag.19.

**Chiaroni, J.** et al. (2008) Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent. *Antiquity* 82, 281–289.

**Chikhi, L.** et al. (1998) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc. Natl. Acad. Sci. U.S.A* 95, 9053–9058.

**Childe, V.G.** (1925) *The dawn of European civilization.* Kegan Paul, London.

**Clutton-Brock, J.** (1999) *A natural history of domesticated animals.* Cambridge University Press, Cambridge.

**Conдеми, S.** et al. (2013) Possible interbreeding in late Italian Neanderthals? New data from the Mezzena jaw (Monti Lessini, Verona, Italy). *PLoS One.* 8(3):e59781.

**Cooper, A., Poinar, H.N.** (2000) Ancient DNA: do it right or not at all. *Science* 289, 1139.

**Dabney, J.** et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15758-15763.

**D'Andria, F.** (a cura di) (1989) *Archologia dei Messapi. Catalogo della Mostra,* Bari 1990.

**D'Andria, F.** (1995) La scoperta delle fortificazioni, della porta urbica ed iscrizioni messapiche del IV-III secolo a.C. Otranto.

**Deguilloux, M.F.** et al. (2011) News from the west: ancient DNA from a French megalithic burial chamber. *Am. J. Phys. Anthropol.* 144, 108–118.



- Denaro, M.** et al. (1981) Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 78, 5768-5772.
- Diamond, J.** (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418:700-707.
- Di Gaetano, C.** et al. (2012) An Overview of the Genetic Structure within the Italian Population from Genome-Wide Data. *PLoS One* 7: e43759.
- Di Giacomo, F.** et al. (2003) Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol. Phylogenet. Evol.* 28, 387–395.
- Di Giacomo, F.** et al. (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum. Genet.* 115, 357–371.
- Edmonson, M.S.** (1961) Neolithic diffusion rates. *Curr. Anthropol.* 2, 71–102.
- Edwards, C.J.** et al (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proc. Biol. Sci.* 274, 1377–1385.
- Edwards, C.J.** et al. (2010) A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS One.* 5(2):e9255.
- Ellegren, H.** (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435-445.
- Ermini, L.** et al. (2008) Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol.* 18, 1687-1693.
- Fix, A.G.** (ed.) (1999) *Migration and Colonization in Human Microevolution.* Cambridge University Press.
- Forster, P.** et al. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59, 935–945.
- Fort, J.** et al. (2012) Modeling the Neolithic transition in the Near East and Europe. *Am. Antiq.* 77, 203–220.

- Francois, O.** et al. (2010) Principal component analysis under population genetic models of range expansion and admixture. *Mol. Biol. Evol.* 27, 1257–1268.
- Fu, Q.** et al. (2013) A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Curr. Biol.* 23, 553-559.
- Fu, Q.** et al (2013b) DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2223-2227.
- Gamba, C.** et al. (2012) Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* 21, 45–56.
- Gamble, C.** et al. (2005) The archaeological and genetic foundations of the European population during the Late Glacial: implications for ‘agricultural thinking’. *Camb. Archaeol. J.* 15, 193–223.
- Gansauge, M.T., Meyer, M.** (2010) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8, 737-748.
- Ghirotto, S.** et al. (2013) Origins and evolution of the Etruscans' mtDNA. *PLoS One.* 8(2):e55519.
- Gilbert, M.T.P.** et al. (2005) Assessing ancient DNA studies. *Trends Ecol. Evol.* 20, 541–544.
- Gilbert, M.T.** et al. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787-1789.
- Gimbutas, M.** (1970) Proto-Indo-European culture: the Kurgan culture during the 5th to the 3rd millennia B.C. En: Cardona G., Koenigswald H.M., Senn A. (Eds.).1970. *Indo-European and Indo-Europeans.* University of Pennsylvania Press. Philadelphia. pp.155-198.
- Green, R. E.** et al. (2008) A complete Neandertal mitochondrial genome sequence determined by highthroughput sequencing. *Cell* 134, 416–426.
- Green, R. E.** et al. (2009) The Neandertal genome and ancient DNA authenticity. *EMBO J.* 28, 2494–2502.
- Green, R.E.** et al. (2010) A draft sequence of the Neandertal genome. *Science* 328, 710–722.

- Guimaraes, S.** et al. (2009) Genealogical discontinuities among Etruscan, Medieval, and contemporary Tuscans. *Mol. Biol. Evol.* 26, 2157-2166.
- Haak, W.** et al. (2010) Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 8(11):e1000536.
- Harris, D.R.** (1996) The origins and spread of agriculture and pastoralism in Eurasia. UCL Press, London.
- Hervella, M.** et al. (2012) Ancient DNA from hunter-gatherer and farmer groups from northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS ONE* 7, e34417.
- Higham, T.** et al. (2011a) The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479, 521–524.
- Higham, T.** (2011b) European Middle and Upper Palaeolithic radiocarbon dates are often older than they look: problems with previous dates and some remedies. *Antiquity* 85, 235–249.
- Higuchi, R.** et al. (1984) DNA-sequences from the Quagga, an extinct member of the horse family. *Nature* 312, 282–284.
- Ho, S.Y.,** Larson, G. (2006) Molecular clocks: when times are a-changin'. *Trends Genet.* 22, 79-83.
- Hoss, M.** et al. (1996) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 24, 1304–1307.
- Hughey, J.R** et al. (2013) A European population in Minoan Bronze Age Crete. *Nat. Commun.* 4:1861.
- Ingman, M.** et al. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708-713.
- Itan, Y.** et al. (2009) The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5, e1000491.
- Jobling, M.A.** et al., eds (2004) *Human Evolutionary Genetics: Origins, Peoples and Disease.* Garland Science.
- Jochim, M.** (2011) The Upper Paleolithic European prehistory. In *European Prehistory: A Survey* (Milisauskas, S., ed.), pp. 67–124, Springer.

- Jonhson, M.J.** et al. (1983) Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* 19, 255-271.
- Keller, A.** et al., (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* 3, 698.
- Kimura, M.** (1955) Random genetic drift in a multiallelic locus. *Evolution* 9, 419-435.
- Knapp, M.** et al. (2012) Complete mitochondrial DNA genome sequences from the first New Zealanders. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18350-18354.
- Krause, J.** et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439, 724–727.
- Krause, J.** et al. (2007) The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* 17, 1908–1912.
- Krause, J.** et al (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.* 8, 220.
- Krause, J.** et al. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, 894-897.
- Kuijt, I., Goring-Morris, N.** (2002) Foraging, farming, and social complexity in the Pre-Pottery Neolithic of the southern Levant: a review and synthesis. *J. World Prehistory* 16, 361–364.
- Lacan, M.** et al. (2011) Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1788–1791.
- Lalueza-Fox, C.** et al. (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-1455.
- Lao, O.** et al. (2008) Correlation between genetic and geographic structure in Europe. *Curr. Biol.* 18, 1241–1248.
- Lari, M.** et al. (2010) The microcephalin ancestral allele in a Neanderthal individual. *PLoS One* 5(5):e10648.

- Lari, M.** et al. (2011) The complete mitochondrial genome of an 11,450-year-old aurochs (*Bos primigenius*) from Central Italy. *BMC Evol. Biol.* 2011 11:32.
- Larson, G.** et al. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc. Natl. Acad. Sci. U.S.A* 107, 7686–7691.
- Larson, G.** (2011) Genetics and domestication: important questions for new answers. *Curr. Anthropol.* 52, S485–S495.
- Leonard, J.A.** et al (2000) Population genetics of ice age brown bears. *Proc. Natl. Acad. Sci. U.S.A* 97, 1651–1654.
- Leonard, J.A.** et al. (2002) Ancient DNA evidence for Old World origin of New World dogs. *Science* 298, 1613–1616.
- Leonard, J.A.** et al (2007) Animal DNA in PCR reagents plagues ancient DNA research. *J. Archaeol. Sci.* 34, 1361–1366.
- Lindahl, T.** (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
- Luikart, G.** et al. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. U.S.A* 98, 5927-5932.
- Lutz, S.** et al. (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int. J. Legal. Med.* 111, 67-77.
- Malmstrom, H.** et al. (2009) Ancient DNA reveals lack of continuity between neolithic huntergatherers and contemporary Scandinavians. *Curr. Biol.* 19, 1758–1762.
- Manzi, G.** et al. (2010) The new chronology of the Ceprano calvarium (Italy). *J. Hum. Evol.* 59, 580-5.
- Maricic, T.** et al. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5, e14004.
- Mellars, P.** (2006) Archaeology and the dispersal of modern humans in Europe: deconstructing the ‘Aurignacian’. *Evol. Anthropol.* 15, 167–182.

- Menozi, P.** et al. (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201, 786–792.
- Miller, W.** et al. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456, 387–390.
- Mona, S.** et al. (2010) Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evol. Biol.* 26, 10:83.
- Morin, E.N.** (2008) Evidence for declines in human population densities during the early Upper Paleolithic in western Europe. *Proc. Natl. Acad. Sci. U.S.A.* 105, 48–53.
- Mounier, A.** et al. (2011) The stem species of our species: a place for the archaic human cranium from Ceprano, Italy. *PLoS One.* 6(4):e18821.
- Naderi, S.** et al. (2008) The goat domestication process inferred from largescale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17659–17664.
- Nei, M.** et al. (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29, 1-10.
- Nelis, M.** et al. (2009) Genetic structure of Europeans: a view from the North East. *PLoS One* 4: e5472.
- Nielsen, R.,** Beaumont, M.A. (2009) Statistical inferences in phylogeography. *Mol. Ecol.* 18, 1034–1047.
- Novembre, J.,** Stephens, M. (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* 40, 646–649.
- Novembre, J.** et al. (2008) Genes mirror geography within Europe. *Nature* 456, 98–101.
- Orlando, L.** et al (2009) Revising the recent evolutionary history of equids using ancient DNA. *Proc. Natl. Acad. Sci. U.S.A* 106, 21754–21759.
- Orlando, L.** et al. (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74-78.
- Ottoni, C.** et al. (2009) Human mitochondrial DNA variation in Southern Italy. *Ann. Hum. Biol.* 36, 785–811.

- Pääbo, S.** (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314, 644–645.
- Pääbo, S.** (1989) Ancient DNA—extraction, characterization, molecular-cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. U.S.A* 86, 1939–1943.
- Pääbo, S.** et al. (1989) Ancient DNA and the polymerase chain-reaction— the emerging field of molecular archaeology. *J. Biol. Chem.* 264, 9709–9712.
- Pääbo, S.** et al. (2004) Genetic analyses from ancient DNA. *Annu. Rev. Genet.* 38, 645–679.
- Pagliara, C.** (1983) Materiali iscritti arcaici del Salento. In *AnnPisa*, s. III-XIII pp. 21-89.
- Pala, M.** et al. (2009) Mitochondrial haplogroup U5b3: a distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *Am. J. Hum. Genet.* 84, 814–821.
- Pala, M.** et al. (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* 90: 915–924.
- Pallottino, M.** (2000) *La Sardegna nuragica*. A cura di Giovanni Lillu, Roma, Edizioni del Gremio (Biblioteca sarda; 53).
- Pavlov, P.** et al. (2001) Human presence in the European Arctic nearly 40,000 years ago. *Nature* 413, 64–67.
- Pavúk, J.** (2005) Typologische Geschichte der Linearbandkeramik. In *Die Bandkeramik im 21. Jahrhundert: Symposium in der Abtei Brauweiler bei Köln* (Lüning, J. et al., eds), pp. 17–39 (in German).
- Pereira, L.** et al. (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15, 19–24.
- Pesando, F.** (2005) *L'Italia antica. Culture e forme del popolamento nel I millennio a. C.* Roma: Carocci editore. 326.
- Pessina, A., Tinè, V.** (2008) *Archeologia del Neolitico. L'Italia tra il VI e il IV millennio a.C.* Roma: Carocci editore. 375.

- Piazza, A.** et al. (1995) Genetics and the origin of European languages. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5836-5840.
- Piercy, R.** et al. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.* 106, 85-90.
- Pilli, E.** et al. (2013) Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. *PLoS One* 8(1):e52524.
- Pinhasi, R.** et al. (2005) Tracing the origin and spread of agriculture in Europe. *PLoS Biol.* 3(12):e410.
- Pinhasi, R.** and von Cramon-Taubadel, N. (2009) Craniometric data support demic diffusion model for the spread of agriculture into Europe. *PLoS ONE* 4, e6747
- .
- Pinhasi, R.** et al. (2011) Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8611–8616.
- Pinhasi, R.**, von Cramon-Taubadel, N. (2012) A craniometric perspective on the transition to agriculture in Europe. *Hum. Biol.* 84, 45–66.
- Pinsky, M.L.** et al (2010) Dispersal provided resilience to range collapse in a marine mammal: insights from the past to inform conservation biology. *Mol. Ecol.* 19, 2418–2429.
- Poinar, H.N.** et al. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311, 392–394.
- Posada, D.** et al. (2000) GeoDis: A program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. *Mol. Ecol.* 9, 487-488.
- Rasmussen, M.** et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757–762.
- Reich, D.** et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060.
- Reich, D.** et al. (2011) Denisova admixture and the first modern human dispersals into southeast Asia and Oceania. *Am. J. Hum. Genet.* 89, 516–528.



- Rohland, N.**, Hofreiter, M. (2007) Ancient DNA extraction from bones and teeth. *Nat. Protoc.* 2, 1756-1762.
- Richards, M.** et al. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* 62 (P3), 241-260.
- Richards, M.** et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67, 1251-1276.
- Rootsi, S.** 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.
- Sanchez-Quinto, F.** et al. (2012) Genomic affinities of two 7,000-year-old Iberian Hunter-Gatherers. *Curr. Biol.* 22, 1494-1499.
- Santachiara-Banerecetti, A.S.** et al. (1988) Mitochondrial DNA polymorphisms in Italy.II. Molecular analysis of new and rare morphs from Sardinia and Rome. *Ann. Hum. Genet.* 52, 39-56.
- Savolainen, P.** et al. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298, 1610-1613.
- Sawyer, S.** et al. (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One.* 7(3):e34131.
- Sboner, A.** et al. (2011) The real cost of sequencing: higher than you think! *Gen. Biol.* 12:125.
- Schuenemann, V.J.** et al. (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341, 179-183.
- Scozzari, R.** et al. (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am. J. Hum. Genet.* 43, 534-544.
- Semino, O.** et al. (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159.
- Shapiro, B.** et al (2002) Flight of the dodo. *Science* 295, 1683.
- Shapiro, B.** et al. (2004) Rise and fall of the Beringian steppe bison. *Science* 306, 1561–1565.

- Sherratt, A.** (1997) Climatic cycles and behavioural revolutions: the emergence of modern humans and the beginning of farming. *Antiquity* 71, 271–287.
- Skoglund, P.** et al. (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336, 466–469.
- Slatkin, M.,** Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555–562.
- Soares, P.** et al. (2010) The archaeogenetics of Europe. *Curr. Biol.* 20, 174–183.
- Sokal, R.R.** (1988) Genetic, geographic, and linguistic distances in Europe. *Proc. Natl. Acad. Sci. U.S.A.* 85, 1722–1726.
- Sokal, R.R.** et al. (1999) A problem with synthetic maps. *Hum. Biol.* 71, 1–13.
- Stewart, J.R.** and Stringer, C.B. (2012) Human evolution out of Africa: the role of refugia and climate change. *Science* 335, 1317–1321.
- Stiller, M.** et al (2010) Withering away-25,000 years of genetic decline preceded cave bear extinction. *Mol. Biol. Evol.* 27, 975–978.
- Stoneking, M.,** Wilson, A.C. (1989) Mitochondrial DNA. In: Hill AVS, Serjeantson S (eds) *The colonization of the Pacific: a genetic trail.* Oxford University Press, Oxford, 215–245.
- Stoneking, M.** et al. (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124, 717–733.
- Stoneking, M.,** Krause, J. (2011) Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet.* 12, 603–14.
- Sykes, B.** et al. (1995) The origins of the Polynesians—an interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* 57, 1463–1475.
- Tambets, K.** et al. (2004) The western and eastern roots of the Saami - the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* 74, 661–682.
- Tassi, F.** et al.(2013) Genetic evidence does not support an Etruscan origin in Anatolia. *Am J. Phys. Anthropol.* 152, 11–8.

**Templeton, A.R.** et al. (1995) Separating population structure from population history: a cladistics analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger salamander, *Ambystoma tigrinum*. *Genetics*. 140, 767-782.

**Thomas, M.G.** et al. (2013) Europe and western Asia: genetics and population history. In *The Encyclopedia of Global Human Migration* (Bellwood, P., ed.), Wiley-Blackwell.

**Torrioni, A.** et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and NaDen populations were founded by two independent migrations. *Genetics* 130: 153-162.

**Torrioni, A.** et al. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53, 563–590.

**Torrioni, A.** et al. (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-Loop region. *Am. J. Hum. Genet.* 55, 760-776.

**Torrioni, A.** et al. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics*. 144, 1835-1850.

**Torrioni, A.** et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* 69, 844–852.

**Torrioni, A.** et al. (2001a) Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates? *Am J Hum Genet.* 69, 1348–1356.

**Torrioni, A.** et al. (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet.* 22, 339–345.

**Turchi, C.** et al. (2008) Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. *Int. J. Legal. Med.* 122, 199–204.

**Underhill, P.A.** and Kivisild, T. (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Ann. Rev. Genet.* 41, 539–564.

**Van Compernelle, T.** (2001) Dall'insediamento iapigio alla città messapica: dieci anni di scavi e ricerche archeologiche a Soletto (LE).

- Verpoorte, A.** (2009) Limiting factors on early modern human dispersals: the human biogeography of late Pleniglacial Europe. *Quatern. Int.* 201, 77–85.
- Vigilant, L.** et al. (1991) African populations and the evolution of human mitochondrial DNA. *Science.* 253, 1503-1507.
- Vigne, J.D.** (2011) The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *C. R. Biol.* 334, 171–181.
- Vilà, C.** et al. (1997) Multiple and ancient origins of the domestic dog. *Science* 276, 1687-1689.
- von Cramon-Taubadel, N.,** Pinhasi, R. (2011) Craniometric data support a mosaic model of demic and cultural Neolithic diffusion to outlying regions of Europe. *Proc. R. Soc. B* 278, 2874–2880.
- Waheley, J.** (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.* 37, 613-623.
- Wang, X.** et al. (2012) Specific inactivation of two immunomodulatory SIGLEC genes during human evolution. *Proc Natl. Acad. Sci. U.S.A* 109, 9935-9940.
- Watterson, G.A.** (1986) The homozygosity test after a change in population size. *Genetics* 112, 899-907.
- Webster, G.S.** (1996) A prehistory of Sardinia 2300–500 BC. SheYeld Academic Press, SheYeld.
- Willerslev, E.,** Cooper, A. (2005) Ancient DNA. *Proc. Biol. Sci.* 272, 3–16.
- Zeder, M.A.** (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11597–11604.
- Zeder, M.A.** (2011) The origins of agriculture in the Near East. *Curr. Anthropol.* 52, S221–S235.

## **APPENDICE 1**

**Tabella A.1:** campioni analizzati, sito di provenienza e datazione. Sono indicate le fasi sperimentali per le quali sono stati ottenuti risultati.

**SITO ARCHEOLOGICO:** Colfiorito (PG)

**DATAZIONE:** IX-V secolo a.C.

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	emPCR	SEQUENZIAMENTO
2	✓	ok	✓ MID3	✓ 1cpb	/
5	✓	ok	✓ MID6	✓ 1cpb	✓
8	✓	ok	✓ MID7	✓ 1cpb	✓
9	✓	ok	✓ MID1	✓ 1cpb	✓
12	✓	ok	✓ MID2	✓ 1cpb	✓
23	✓	ok	✓ MID5	✓ 1cpb	✓
23N	✓	ok	✓ MID6	✓ 1cpb	✓
35	✓	ok	✓ MID7	✓ 1cpb	✓
39	✓	ok	✓ MID8	✓ 1cpb	✓
41	✓	/	✓	✓	✓
42	✓	ok	✓ MID2	✓ 1cpb	✓
44	✓	ok	✓ MID11	✓ 1cpb	✓
46	✓	ok	✓ MID12	✓ 1cpb	✓
49	✓	ok	✓ MID3	✓ 1cpb	✓
50	✓	ok	✓ MID8	✓ 1cpb	✓
53	✓	ok	✓ MID10	✓ 1cpb	✓
55	✓	ok	✓ MID4	✓ 1cpb	✓

58	✓	ok	✓ MID8	✓ 1cpb	✓
59	✓	ok	✓ MID6	✓ 1cpb	/
61	✓	ok	✓ MID7	✓ 1cpb	/
62	✓	ok	✓ MID2	✓ 1cpb	✓
73	✓	ok	✓ MID5	✓ 1cpb	✓
79	✓	ok	✓ MID1	✓ 1cpb	✓
82	✓	I, II tr	/	/	/
109	✓	ok	✓ MID2	✓ 0.5cpb	✓
125	✓	II, III tr	/	/	/
133	✓	ok	✓ MID11	✓ 0.5cpb	✓
137	✓	ok	✓ MID8	✓ 1cpb	✓
138	✓	I, III tr	/	/	/
141	✓	ok	✓ MID6	✓ 1cpb	✓
142	✓	ok	✓ MID5	✓ 1cpb	✓
144	✓	ok	✓ MID11	✓ 1cpb	✓
147	✓	ok	✓ MID4	✓ 1cpb	✓
151	✓	ok	✓ MID1	✓ 1cpb	✓
155	✓	ok	✓ MID4	✓ 1cpb	/
192	✓	ok	✓ MID8	✓ 1cpb	✓
198	✓	ok	✓ MID12	✓ 1cpb	✓
204	✓	I, III tr	/	/	/

209	✓	ok	✓	MID10	✓	1cpb	✓
211		ok					
222	✓	ok	✓	MID1	✓	1cpb	✓
229	✓	ok	✓	MID5	✓	1cpb	✓
231	✓	ok	✓	MID8	✓	1cpb	✓
233	✓	III tr		/		/	/
294	✓	ok	✓	MID2	✓	1cpb	✓



**SITO ARCHEOLOGICO: Salve (LE)****DATAZIONE: 2.500-2.300 a.C.**

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	emPCR	SEQUENZIAMENTO
<b>4L7</b>	✓	ok	✓ MID7	✓ 1cpb	✓
<b>40L5</b>	✓	ok	✓ MID11	✓ 0.5cpb	✓
<b>68L7</b>	✓	ok	✓ MID4	✓ 1cpb	✓
<b>75L6</b>	✓	ok	✓ MID4	✓ 1cpb	✓
<b>11L7</b>	✓	ok	✓ MID1	✓ 1cpb	✓

**SITO ARCHEOLOGICO: Roca Vecchia (LE)****DATAZIONE: IV-III sec a.C.**

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	emPCR	SEQUENZIAMENTO
<b>1-T16</b>	✓	II tr	/	/	/
<b>3-T4</b>	✓	/	/	/	/
<b>4-T5</b>	✓	II tr	/	/	/
<b>5-T6</b>	✓	ok	✓ MID6	✓ 1cpb	/
<b>7-T6</b>	✓	I tr	/	/	/
<b>8-T19</b>	✓	II, III tr	/	/	/
<b>10-T7</b>	✓	ok	✓ MID11	✓ 0.5cpb	✓

11-T7	✓	III tr	/	/	/
12-T7	✓	I, II tr	/	/	/
13-T7	✓	ok	✓ MID5	✓ 1cpb	✓
17-T20	✓	II tr	/	/	/
20-T4/3	✓	II, III tr	/	/	/

**SITO ARCHEOLOGICO:** Mesagne (BR)

**DATAZIONE:** IV-III secolo a.C.

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	emPCR	SEQUENZIAMENTO
231-T8	✓	II, III tr	/	/	/
232-T4	✓	ok	✓ MID6	✓ 1cpb	✓
233-T5	✓	II, III tr	/	/	/
234-T7	✓	I, III tr	/	/	/
235-T4/90	✓	I tr	/	/	/
236-T5	✓	ok	✓ MID10	✓ 0.5cpb	✓
T1 (SP88)	✓	/	/	/	/
T2 (SP)	✓	I, III tr	/	/	/

<b>T6 ('88)</b>	✓	Il tr	/	/	/
<b>T6 ('91)</b>	✓	/	/	/	/
<b>TA</b>	✓	/	/	/	/
<b>D1 ('91)</b>	✓	/	/	/	/
<b>D2 ('91)</b>	✓	/	/	/	/
<b>92/1</b>	✓	/	/	/	/

**SITO ARCHEOLOGICO: Su Forru de is Sinzurreddus (OR)**

DATAZIONE: 4.000-3.400 a.C.

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	EmPCR	SEQUENZIAMENTO
SZ91	✓	II tr	/	/	/
SZ92	✓	ok	✓ MID3	✓ 1cpb	✓
SZ93	✓	ok	✓ MID11	✓ 1cpb	✓
SZ94	✓	/	/	/	/
SZ114	✓	/	/	/	/
SZ123	✓	III tr	/	/	/

**SITO ARCHEOLOGICO: Cuccuru is Arrius (OR)**

DATAZIONE: V-IV millennio a.C.

CAMPIONE	ESTRAZIONE	HVR1	LIBRARY/MID	EmPCR	SEQUENZIAMENTO
CA8	✓	/	/	/	/
CA9	✓	II tr	/	/	/
CA10	✓	II tr	/	/	/
CA11	✓	II tr	/	/	/
CA12	✓	III tr	/	/	/

<b>CA13</b>	✓	/	/	/	/
<b>CA20</b>	✓	II, III tr	/	/	/
<b>CA21</b>	✓	II tr	/	/	/
<b>CA22</b>	✓	II, III tr	/	/	/
<b>CA23</b>	✓	/	/	/	/

**SITO ARCHEOLOGICO: Siamaggiore (OR)**

**DATAZIONE: 3.200-2.800 a.C.**

CAMPIONE	ESTRAZIONE	HVRI	LIBRARY/MID	EmPCR	SEQUENZIAMENTO
<b>SI01</b>	✓	ok	✓ MID5	✓ 1cpb	✓
<b>SI13</b>	✓	I, II tr	/	/	/

**SITO ARCHEOLOGICO: Su Carroppu di Sirri (OR)**

DATAZIONE: 6.000-4.0000 a.C.

CAMPIONE	ESTRAZIONE	HVRI	LIBRARY/MID	EmPCR	SEQUENZIAMENTO
CAR_H10	✓	ok	✓ MID8	✓ 1cpb	/
CAR_H12	✓				
CAR_H13	✓	ok	✓ MID2	✓ 1cpb	✓
CAR_H14	✓	I, II tr	/	/	/
CAR_H15	✓	II, III tr	/	/	/
CAR_H18	✓	ok	✓ MID3	✓ 1cpb	✓
CAR_H19	✓	ok	✓ MID7	✓ 1cpb	✓
CAR_H20	✓	/	/	/	/
CAR_H21	✓	II, III tr	/	/	/

**Tabella A.2:** aplotipi e linee mitocondriali di afferenza dei campioni analizzati.  
h = eteroplasmia  
d = delezione

a) UMBRI

CAMPIONE	Hp	Hg	QUALITA'	OSSERVAZIONI
5	16311h	H	ok	Mixed?
8	16298, 16086	V/Ub31a	ok	Mixed?
9	16042, 16126, 16304, 16224h	T2	ok	Possibile
12	16069, 16126, 16294, 16256, 16234h	J1c	ok	Possibile
23	16147A, 16172, 16223, 16246	N1a	ok	Perfetto
23N	16147	H	ok	Possibile
35	16126, 16183C, 16256, 16294, 16324	T2	ok	Perfetto
39	16223	W	??	Possibile
42	16188G, 16115h, 16184h	H	ok	Possibile
44	16293, 16311, 16205h, 16224h	K	ok	Possibile
46	16290, 16328, 16337, 16393, 16224h	??	no	Mixed
49	16114A, 16126, 16284, 16184h	??	no	Mixed
50	16126, 16208, 16069h	J1c	ok	Possibile
53	16189, 16291, 16357, 16182Ch, 16183Ch	H	ok	Perfetto
55	16069, 16126, 16390, 16251h, 16296h	J1c	ok	Buono
58	16124, 16261, 16289, 16290	??	??	Mixed
62	16124, 16234	K	ok	Possibile
73	16184, 16294, 16296, 16304	T2b	ok	Possibile/Mixed con Umbri 79?
79	16126, 16183C, 16189, 16293, 16311, 16375h	R0a	ok	Possibile

<b>109</b>	16124, 16256, 16270, 16273, 16291, 16384, 16388, 16389, 16126h	??	no	Mixed
<b>133</b>	16114, 16164, 16183C, 16189, 16256	U5a	ok	Perfetto (manca III tr)
<b>137</b>	16069, 16126, 16150, 16183C, 16189, 16304	J1	ok	Possibile
<b>141</b>	16164h, 16183C, 16189, 16375A	H	ok	Possibile
<b>142</b>	16294, 16304, 16320, 16335, 16126h	T2b	ok	Possibile
<b>144</b>	16224, 16234, 16070h	K1	ok	Perfetto (manca III tr)
<b>147</b>	16293, 16311, 16375Ah, 16259delh?	H	ok	Possibile
<b>151</b>	16114, 16126, 16193, 16259, 16263, 16264, 16278, 16278, 16284, 16288, 16290, 16293C, 16301, 16311, 16355, 16356, 16368, 16390	L1b/X?	??	Mixed
<b>155</b>	16183C, 16189, 16304, 16164h	H	ok	Perfetto
<b>192</b>	16051, 16092, 16164, 16293h, 16311h	U2	??	Possibile
<b>198</b>	16067, 16079, 16164, 16293, 16311	HV1	??	??
<b>209</b>	16129, 16298, 16164h, 16311h	V	ok	Possibile
<b>211</b>	16051, 16069, 16260, 16278, 16294, 16309, 16328, 16390, 16375Ah	??	no	Mixed
<b>222</b>	16039, 16042, 16047, 16048, 16049, 16084, 16223, 16256, 16270, 16294	??	no	Mixed
<b>229</b>	16126, 16183C, 16189, 16256A, 16324, 16294h	T2a	ok	Possibile
<b>231</b>	16290, 16298, 16377	V	ok	Possibile
<b>294</b>	16192, 16256, 16293, 16311	U5a/H	??	192 e 259 = U5a 293 e 311 = H



b) **MESSAPI**

CAMPIONE	Hp	Hg	QUALITA'	OSSERVAZIONI
<b>10T7</b>	16129, 16223, 16311	I	ok	Perfetto
<b>11T7</b>	16182C, 16183C, 16187, 16189, 16223	??	??	Mixed
<b>13T7</b>	16126	H	ok	Possibile
<b>40L5</b>	16166d, 16261	H	ok	Perfetto
<b>68L7</b>	16126, 16295h	H	ok	Possibile
<b>75L6</b>	16261, 16291	H	ok	Perfetto
<b>232T4</b>	16183C, 16189, 16293, 16311	H	ok	Perfetto
<b>236T5</b>	16223, 16278, 16390	L2	ok	Mai trovato in Europa
<b>4L7</b>	16093, 16189, 16390	H	ok	Perfetto

c) **SARDI**

CAMPIONE	Hp	Hg	QUALITA'	OSSERVAZIONI
<b>CAR_H13</b>	16224, 16311	K	ok	Perfetto
<b>CAR_H18</b>	16126, 16294, 16296	T2	ok	Perfetto
<b>CAR_H19</b>	16265, 16270, 16294, 16295	??	??	Mixed
<b>SI01</b>	16224, 16311, 16293	K	ok	Perfetto
<b>SZ92</b>	16166d, 16360	H	ok	Molto raro in questa forma
<b>SZ93</b>	CRS	H	ok	Perfetto (copertura non molto buona)

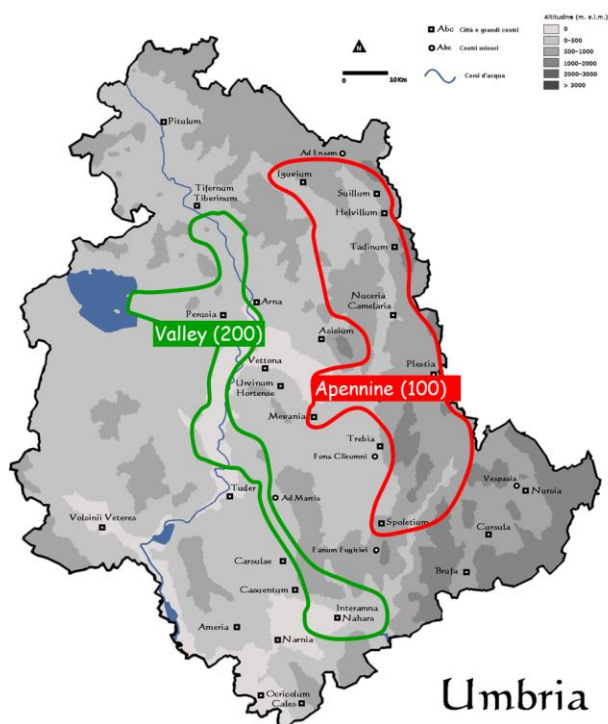
## **APPENDICE 2**

## UMBRI MODERNI

Tutte le analisi sui campioni moderni sono state svolte presso l'Università di Perugia.

### 1.A2 CAMPIONAMENTO

Sono stati raccolti 300 campioni biologici – mediante buccal swabs – da partecipanti volontari provenienti da due diverse aree dell'Umbria (Fig. 1.A2). Il primo campione comprende 100 individui dell'altopiano di Colfiorito e dei paesi vicini (Appennino centrale); mentre il secondo campione consta di 200 individui provenienti dalla regione di valle, compresa Perugia.



**Fig. 1.A2:** aree dalle quali provengono i campioni umbri moderni: in rosso il campione appenninico; in verde quello di valle.

Il campionamento è stato effettuato con l'aiuto di assistenti locali. Ad ogni soggetto partecipante è stato presentato e fatto firmare un modulo di consenso informato (Autorizzazione al trattamento dei dati genetici del 22 febbraio 2007, da parte del Garante per la Protezione dei Dati Personali, Gazzetta Ufficiale n° 65 del 19 marzo 2007). La genealogia dei soggetti è stata identificata mediante un albero genealogico con l'attenzione rivolta agli antenati materni, assicurandoci della loro provenienza da borghi dell'Umbria. Ogni campione è stato quindi siglato con un numero di codice progressivo per la tutela dell'anonimato.

## **2.A2 METODI**

a) Estrazione del DNA. Il DNA è stato estratto con kit commerciale Wizard Genomic DNA Purification Kit (Promega) secondo il manuale fornito dalla casa produttrice.

Una volta estratti, i DNA sono stati sottoposti a determinazione quantitativa mediante spettrofotometro.

b) Amplificazione della regione di controllo. Il primo e il secondo segmento ipervariabile (HVR-1 e HVR-2) sono stati amplificati mediante PCR. Per ciascuna amplificazione sono stati utilizzati circa 2 ng di template in una miscela di reazione così costituita: 1X Reaction Buffer (Promega), 100 mM di ciascun dNTP, 0.3  $\mu$ M di ciascun primer (Tabella Umbri\_mod1), 0.04U/ $\mu$ l di Taq Polimerasi (Promega GOTaq®) in un volume finale di 25  $\mu$ l. La reazione è stata svolta con il seguente profilo termico:

- 95°C per 2 minuti
- ripetizione del seguente ciclo per 35 volte:
  - 95°C per 30 secondi
  - 55°C per 30 secondi
  - 72°C per 120 sec + 1 sec./cycle
- 72°C per 5 minuti
- 4°C forever

<i>Nome</i>	<i>Sequenza</i>	<i>Lunghezza</i>
15877For	CAAATGGGCCTGTCCTTGTA	20
250Rev	TGGCAGAGATGTGTTTAAGTGC	22

**Tabella Umbri\_mod1:** primers utilizzati per l'amplificazione della regione di controllo.

c) Sequenziamento degli ampliconi. Dopo purificazione enzimatica (Exosap, USB), a 100 ng di DNA amplificato sono stati aggiunti 3.2  $\mu$ l (2 pmol/ $\mu$ l) degli oligonucleotidi in tabella 2.A2, per il sequenziamento. Il sequenziamento è stato effettuato utilizzando il primer forward, ma in alcuni casi – come nel caso dello stretch di poli-C tra le posizioni nucleotidiche 16184 e 16193 – è stato utilizzato anche il primer reverse.

<i>Nome</i>	<i>Sequenza</i>	<i>Lunghezza</i>
15980For	CCATTAGCACCCAAAGCTA	19
259Rev	TGTGCAGACATTCAATTGTT	20
270Rev	TGATGCTTGTTCTTTTGAT	20

**Tabella Umbri\_mod2:** primers utilizzati per il sequenziamento dell'intera D-loop.

I campioni sono stati quindi inviati alla BMR-Genomics s.r.l (<http://www.bmr-genomics.com/>), per il sequenziamento – *chain-termination dideoxynucleotide* (ddNTPs).

d) Analisi delle sequenze. Le sequenze ottenute sono state assemblate, allineate e confrontate con il software Sequencer™ 4.7 (Gene Codes).

Le relazioni evolutive tra le sequenze ottenute sono state quindi stimate attraverso la costruzione di un network generato con il programma Network 6.1.

### 3.A.2 RISULTATI

Dei 300 individui iniziali, è stata ottenuta la sequenza di 225 campioni, a partire dalla posizione nucleotidica 16000, per un totale di 800 bp, includendo l'intera HVR-1 e parte della regione ipervariabile 2. Sulla base dei polimorfismi presenti abbiamo identificato 177 aplotipi, evidenziando un'elevata diversità genetica ( $Hd = 0.993 \pm 0.002$ ) della popolazione umbra. Attraverso l'analisi dei marcatori diagnostici presenti nella regione codificante, abbiamo classificato i mtDNA in 28 aplogruppi e 18 sotto-aplogruppi.

Come si può osservare dal network – basato sulla regione di controllo e sui polimorfismi della regione codificante – (Fig. 2.A2), i due gruppi analizzati, quello Appenninico e quello di Valle, condividono la maggior parte degli aplogruppi. Circa la metà sono classificati come H (Fig. 3.A2), che è comunque l'aplogruppo più comune in Eurasia, con frequenze simili a quelle riportate per altre regioni d'Europa. Tutti gli altri mtDNA sono quelli tipici dell'Eurasia.

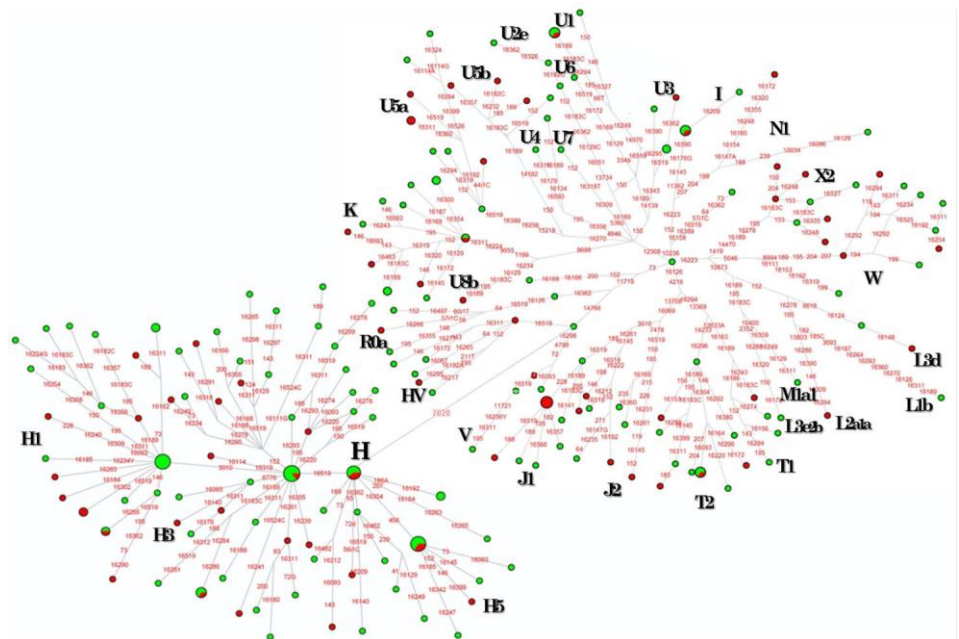
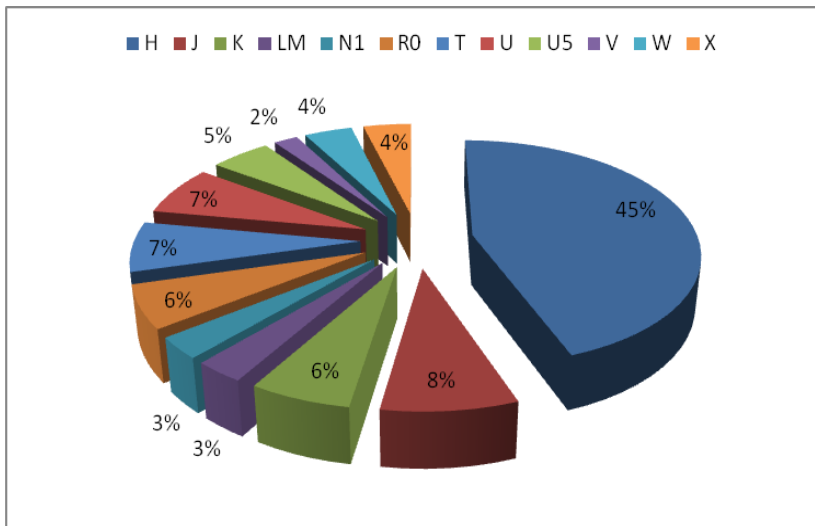


Fig. 2.A2: Median-Joining Network degli umbri moderni. In rosso gli individui appenninici; in verde quelli di valle.



**Fig 3.A2:** frequenze dei principali aplogruppi mitocondriali ritrovati nella popolazione umbra moderna.

Dal network (Fig. 2.A2) emerge inoltre una distribuzione piuttosto omogenea delle frequenze degli aplogruppi tra il gruppo di Valle e quello Appenninico, e il chi-square test esclude la presenza di differenze statisticamente significative tra le due aree geografiche (Pearson  $\chi^2$  18.6; df: 27;  $P$ : 0.88), pertanto l'intero dataset può essere considerato come un'unica variabile.

- **Confronto con le popolazioni dell'Eurasia Occidentale.** Per collocare geneticamente la popolazione umbra presa in esame rispetto alle altre popolazioni sia italiane che dell'Eurasia occidentale siamo ricorsi all'analisi delle componenti principali (PCA). Sono state prese in esame 57 popolazioni mediorientali ed europee, ottenendo il risultato di Figura 4.A2.



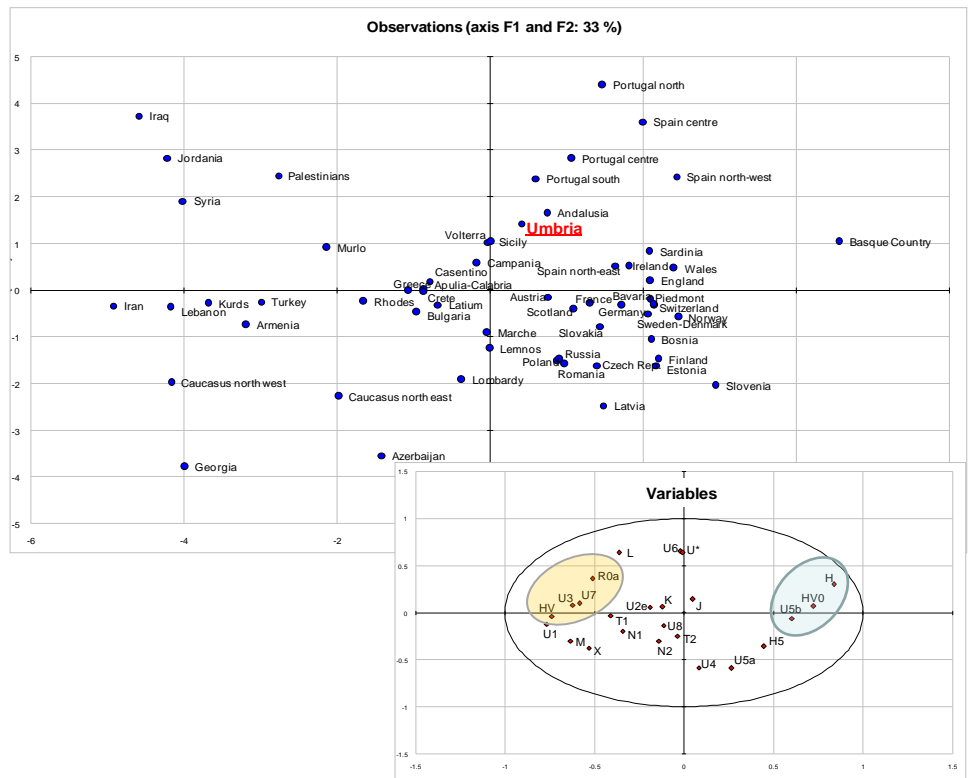


Fig 4.A2: analisi PCA delle popolazioni dell'Eurasia occidentale

Risulta evidente la vicinanza tra gli Umbri e le popolazioni della penisola iberica, supportando l'idea di una vicinanza genetica con l'Europa occidentale piuttosto che con il Medio Oriente.



## **APPENDICE 3**

**Tabella A.3:** popolazioni moderne utilizzare per i confronti inter-popolazionistici

<b>Sigla</b>	<b>Popolazione</b>	<b>N° individui</b>	<b>Ref. Bibliografica</b>
Umbri	Umbri moderni	270	
Aust	Austrians	117	Handt et al. 1994, Parson et al. 1998
Basq	Basques	106	Bertranpetit et al. 1996, Corte-Real et al. 1996
Belg	Belgium	33	De Corte et al. 1996
Brit	British	100	Richards et al. 1996, Piercy et al. 1993
Bulg	Bulgarians	30	Calafell et al. 1996
Cata	Catalans	15	Corte-Real et al. 1996
Chec	Chechenians	23	Nasdize & Stoneking 2001
Corn	Cornwell	69	Richards et al. 1996
Dane	Danesh	62	Richards et al. 1996
Druz	Druze	45	Macaulay et al.
Esto	Estonians	28	Sajantila et al. 1995
Finl	Finland	79	Sajantila et al. 1995, Richards et al. 1996
Fren	French	171	M. Le Roux (p.c.); Rousselet et al. IntJLegalMed 111:292-298, 1998
Gale	Galecians	92	Salas et al. 1998
Geor	Georgians	102	Nasdize & Stoneking 2001, Comas et al., 2000
Gern	Germans north	108	Richards et al. 1996
Gers	Germans south	249	Richards et al. 1996, Lutz et al. 1998
Ingu	Ingushians	35	Nasdize & Stoneking 2001
Kare	Karelians	83	Sajantila et al. 1995
Kurd	Kurds	29	Comas et al
MEast	MEast	42	Di Rienzo and Wilson 1991
Norw	Norwegians	30	Dupuy and Olaisen 1996
Neth	Netherlands	166	provided by Peter de Knijff
Port	Portugal	54	Corte-Real et al. 1996
Russ	Russians	241	Belayeva et al. 2003; Malyarchuk et al 2001; Sajantila et al. 1995, Dupuy and Olaisen 1996
Spap	Spain	108	Casas et al., 2006
Swed	Sweden	32	Sajantila et al. 1995; Dupuy and Olaisen, 1996
Swid	Swiss	74	Pult et al. 1994
Siri	Sirians	49	Vernesi et al., 2001
Volg	Volga-Finnic	34	Sajantila et al. 1995

Wale	Wales	92	Richards et al. 1996, Piercy et al. 1993
Bologna	Bologna	100	Bini et al 2003
Campania	Campania	48	Babalini et al 2005
Lazio	Lazio	52	Babalini et al 2005
Albania	Albania	42	Belledi
Molise	Molise	62	Babalini et al 2005
Abruzzo	Abruzzo	11	Babalini et al 2005
Puglia	Puglia	26	Babalini et al 2005
Grecia	Grecia	52	Kouvatsi
Piem_Torino	Piemontesi Torino	50	Turchi et al. 2008
Murlo	Murlo	68	Achilli
Casentino	Casentino	197	Achilli
Volterra	Volterra	57	Achilli
Tuscan	Toscani Francalacci	49	Francalacci 1996
Turchi	Turchi	35	Di Benedetto
Sardi Ogliastra	Sardi Ogliastra	175	Fraumene et al. 2003
Sardi Gallura	Sardi Gallura	27	Morelli et al. 2000

**Tabella A.4:** popolazioni antiche utilizzare per i confronti inter-popolazionistici

<b>Popolazione</b>	<b>N° individui</b>	<b>Ref. Bibliografica</b>
Etruschi	30	Vernesi, Caramelli et al. 2004; Ghirotto et al. 2013
Messapi	9	Presente lavoro
Sardi	23	Caramelli, Vernesi et al. 2007
Piemontesi	3	Vai et al (Submitted)