

**UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II'**

**DOTTORATO DI RICERCA IN BIOLOGIA AVANZATA**

**XVIII CICLO**

**CARATTERIZZAZIONE DELLA STRUTTURA GENETICA DELLA  
POPOLAZIONE DI CAMPORA**

**Relatore**

**Prof. Gaetano Odierna**

**Candidata**

**Vincenza Colonna**

**Correlatore**

**Dr. Graziella Persico**

**Anno Accademico 2004/2005**

<b>1 - INTRODUZIONE</b>	<b>4</b>
<b>1.1 - Le popolazioni isolate e i meccanismi dell'isolamento</b>	<b>4</b>
<b>1.2 - Studio della struttura delle popolazioni</b>	<b>6</b>
1.2.1 - Inbreeding	6
1.2.2 - Linkage disequilibrium	7
1.2.3 - Struttura delle popolazioni isolate	8
<b>1.3 - I tratti complessi</b>	<b>9</b>
1.3.1 - Utilizzo delle popolazioni isolate per lo studio dei tratti complessi	10
<b>1.4 La popolazione di Campora</b>	<b>11</b>
1.4.1 - Cenni storici	11
1.4.2 - La genealogia di Campora	13
<b>1.5 Scopo del lavoro di tesi</b>	<b>13</b>
<b>2 - RISULTATI</b>	<b>15</b>
<b>2.1 - Notizie di demografia</b>	<b>15</b>
2.1.1 Andamento della popolazione	15
2.1.2 Esogamia	15
<b>2.2 - Analisi del pedigree</b>	<b>16</b>
<b>2.2.1 Linee genealogiche</b>	<b>16</b>
2.2.2 Definizione della popolazione nativa	16
2.2.3 Inbreeding nella popolazione nativa	17
2.2.4 Distribuzione degli individui nelle linee genealogiche	17
<b>2.3 - Determinazione del numero di fondatori della popolazione di Campora</b>	<b>18</b>
2.3.1 Determinazione del numero di linee fondatrici femminili	18
2.3.2 Determinazione del numero di linee fondatrici maschili	20
<b>2.4 Haplotype Diversity sul cromosoma Y</b>	<b>21</b>
<b>2.5 Analisi del PATTERN di Linkage Disequilibrium</b>	<b>22</b>
2.5.1 LD nelle regioni Xq13.3Xq21.1	22
2.5.2 Linkage Disequilibrium nel genoma	23
<b>3- MATERIALI E METODI</b>	<b>24</b>
<b>3.1 - Database</b>	<b>24</b>
<b>3.2 - Ricostruzione delle linee genealogiche</b>	<b>24</b>
<b>3.3 - Campionamento e preparazione di DNA</b>	<b>25</b>
<b>3.4 - Files formato linkage</b>	<b>25</b>

<b>3.5 - Coefficiente di inbreeding</b>	<b>25</b>
<b>3.6 - Linkage Disequilibrium nel genoma</b>	<b>26</b>
3.6.1 - Messa in fase dei marcatori	26
3.6.2 - Calcolo del LD per ciascun cromosoma	27
3.6.3 - Correzione per test multipli	27
3.6.4 - Linkage Disequilibrium sul cromosoma X	27
<b>3.7 - Analisi del DNA mitocondriale</b>	<b>28</b>
<b>3.8 - Analisi del cromosoma Y</b>	<b>29</b>
<b>4 - DISCUSSIONE</b>	<b>30</b>
<b>4.1 - Analisi in Campora delle componenti che caratterizzano le popolazioni isolate</b>	<b>30</b>
4.1.1 - Il bottleneck causato dalla peste	30
4.1.2 - Le linee fondatrici	32
4.1.3 - L'effetto della deriva genetica	33
4.1.4 - La popolazione di Campora è una popolazione inbred	33
4.1.5 - Ampiezza delle regioni in Linkage Disequilibrium	34
<b>4.2 - Considerazioni conclusive sulla popolazione di Campora</b>	<b>35</b>
<b>4.3 - Quanto la genetica può dirci della storia di una popolazione?</b>	<b>36</b>
<b>Bibliografia</b>	<b>37</b>

# 1 - INTRODUZIONE

## 1.1 - Le popolazioni isolate e i meccanismi dell'isolamento

Si definiscono isolate quelle popolazioni che derivano da un ridotto numero di fondatori e persistono in uno stato di isolamento geografico o culturale per un certo numero di generazioni durante le quali non ricevono alcun contributo genetico dall'esterno (Arcos-Burgos and Muenke 2002).

L'effetto fondatore interviene nella fase iniziale dell'isolamento ed è stato definito da Ernst Mayr come “la formazione di una popolazione nuova a partire da pochi fondatori, portatori solo di una piccola frazione della variabilità genetica della popolazione parentale” (Mayr 1976). Una popolazione può discendere da un ridotto numero di fondatori principalmente per due ragioni: (i) un ridotto numero di individui colonizza un posto precedentemente disabitato (ii) si verificano delle fluttuazioni nelle dimensioni della popolazione. Un esempio della prima condizione è il caso dell'isola di Tristan da Cunha (Soodyall et al. 2003), che si trova nell'oceano Atlantico a Sud, equamente distante dall'Africa e dal Sud-America. Esiste un'attendibile documentazione della genealogia di questa popolazione che risale all'epoca del primo insediamento, nel 1816, attestante la discendenza della popolazione attuale da sette femmine e otto maschi. Tale evidenza è stata confermata sperimentalmente (Soodyall et al. 1997; Soodyall et al. 2003). La seconda condizione si verifica invece quando una popolazione passa attraverso un “collo di bottiglia”. Un collo di bottiglia consiste in una repentina riduzione del numero di individui in una popolazione in seguito alla quale i pochi individui sopravvissuti si espandono, in presenza di condizioni favorevoli (**fig1**).

L'isolamento può avere diverse cause, la più importante delle quali è geografica. Le barriere geografiche, in particolare montagne e corsi d'acqua, tendono a ridurre la

mobilità degli individui e quindi a creare i presupposti per l'isolamento. Altre cause possono essere cambiamenti ambientali improvvisi ed il verificarsi di circostanze estreme (carestie, guerre, pestilenze), fattori religiosi o culturali. In **tab1** sono illustrati alcuni esempi di popolazioni umane isolate e per ciascuna sono indicate le cause principali dell'isolamento.

L'isolamento porta alla formazione di un pool genico differente rispetto a quello della popolazione parentale e delle popolazioni circostanti. Il differenziamento del pool genico avviene principalmente a causa dell'adattamento selettivo e della deriva genetica (in assenza di migrazione). L'adattamento selettivo è il processo di risposta degli individui alla pressione esercitata dalla selezione naturale. L'adattamento selettivo dipende dal contesto ambientale. La deriva genetica consiste nella fluttuazione delle frequenze alleliche da una generazione all'altra, dovuta al campionamento casuale dei gameti (**fig2**). La deriva genetica agisce in misura maggiore sulle popolazioni piccole e porta all'estinzione di alcuni alleli e alla fissazione di altri (**fig2**), mentre nelle popolazioni grandi o in rapida espansione è parzialmente compensata dall'elevato numero di progenie per generazione (Cavalli-Sforza et al. 1996).

L'effetto dell'adattamento selettivo e della deriva genetica sono così peculiari che due popolazioni isolate, derivanti dalla stessa popolazione parentale ed aventi quindi, inizialmente, la stessa composizione genetica, possono sviluppare nel corso delle generazioni due differenti pools genici. Tale evento può avvenire se le due popolazioni occupano due ambienti diversi, nei quali subiscono una pressione selettiva differente. Anche occupando lo stesso ambiente, due popolazioni possono sviluppare meccanismi diversi di adattamento allo stesso stimolo selettivo a causa della deriva genetica. Infine va considerato l'effetto dell'accumulo di mutazioni, la cui entità è pressoché costante in tutte le popolazioni, ma i cui risultati su ciascuna popolazione sono completamente differenti.

## 1.2 - Studio della struttura delle popolazioni

Determinare la struttura genetica di una popolazione vuol dire comparare il suo pool genico rispetto alla popolazione da cui deriva o rispetto ad altre popolazioni o ancora rispetto ad un modello nullo in cui si hanno dimensioni della popolazione sufficientemente grandi, in teoria infinite, accoppiamento casuale, frequenze alleliche identiche in maschi e femmine, generazioni sincrone, assenza di mutazioni ed assenza di migrazione.

La comparazione viene effettuata attraverso l'analisi della distribuzione delle frequenze alleliche e lo studio degli aplotipi (combinazioni di alleli a più loci). Questi sono i parametri che meglio descrivono l'effetto dei fattori (deriva genetica, inbreeding, migrazioni, collo di bottiglia, effetto fondatore) che influenzano la composizione del pool genico di una popolazione.

### 1.2.1 - Inbreeding

In una popolazione aperta gli accoppiamenti tra due individui avvengono in modo casuale. In una popolazione isolata, gli accoppiamenti avvengono tra individui che sono legati da relazioni di parentela più o meno strette (**fig3**). Tali individui quindi presentano uno o più antenati in comune. La progenie derivante da accoppiamenti che avvengono in una popolazione isolata, è una popolazione inbred.

Il livello di inbreeding presente in una popolazione viene determinato attraverso il calcolo del coefficiente di inbreeding, definito come la probabilità che in un individuo due alleli a due loci siano identici per discendenza (IBD, Identical By Descendant) da un comune antenato ossia siano in una condizione di autozigosi (al contrario si definiscono allozigoti due alleli in omozigosi identici per stato).

Il coefficiente di inbreeding (F) misura il grado di autozigosità rispetto alla popolazione ancestrale, descrivendo la variazione delle frequenze alleliche in presenza di inbreeding nella formula di Hardy-Weinberg. Tale legge descrive le frequenze dei vari genotipi (omozigoti o eterozigoti per un locus con due alleli) possibili in una

popolazione, a partire dalle frequenze alleliche (**fig4**). Proprio da questa formula  $F$  è stato estrapolato nella sua prima espressione da Sewall Wright nel 1922. In questo modo  $F$  diventa una misura della riduzione di eterozigosità rispetto ad una popolazione panmittica.

Il coefficiente di inbreeding si calcola in modo più esatto utilizzando la genealogia della popolazione (Crow and Kimura 1970). In questo caso si tiene conto del numero di step meiotici che occorre percorrere a partire da un individuo per ritornarvi, passando per i genitori e gli antenati in comune (**fig3**)

L'inbreeding aumenta la variazione fenotipica aumentando l'omozigosità e quindi la possibilità per i caratteri recessivi di manifestarsi. L'inbreeding può quindi causare la segregazione di un carico di genotipi recessivi deleteri, che vengono persi nel lungo periodo poiché conferiscono una ridotta fitness agli individui che li possiedono. In presenza di inbreeding si ha una riduzione della diversità allelica, a causa del modesto pool genico degli individui componenti la popolazione. Questa condizione viene accentuata in caso di assenza di flusso genico (assenza di immigrazioni).

### 1.2.2 - Linkage disequilibrium

Il linkage disequilibrium (LD) o associazione allelica, è il fenomeno per cui due alleli a due loci adiacenti segregano insieme più frequentemente di quanto atteso se fossero indipendenti, condizione nella quale i due loci si dicono in linkage equilibrium (Hartl and Clark 1989). Questo concetto è formalizzato in una delle prime misure proposte per il LD (Lewontin 1964), che lo definisce come la differenza tra la frequenza osservata di un aplotipo a due loci e quella attesa nel caso in cui i due loci segreghino indipendentemente.

Mutazioni e ricombinazioni sono i principali fattori che rispettivamente generano e distruggono il LD (**fig5**). Quando si verifica una mutazione, si genera LD tra il sito mutato e gli alleli adiacenti. Inizialmente la mutazione sarà associata solo a quei particolari alleli, fino a quando non interviene la ricombinazione che separa la mutazione dagli alleli a cui era associata e la congiunge ad altre forme alleliche (Ardlie et al. 2002).

Il livello di LD decresce con il tempo e con la distanza tra i marcatori (frazione di ricombinazione) secondo la legge  $D_t = (1 - r)^t D_0$  dove  $D_0$  rappresenta il LD al momento di inizio e  $D_t$  è il LD dopo  $t$  generazioni. Tuttavia questa equazione non descrive in modo appropriato il LD a distanze piccole per cui può accadere che un polimorfismo non sia in disequilibrio con un marcatore vicino mentre lo sia con un altro che si trova a maggiore, talvolta notevole, distanza. Questo fa capire come i fattori che regolano il LD siano numerosi e complessi: oltre a mutazione e ricombinazione, molti altri fattori, tra cui la deriva genetica e le migrazioni, contribuiscono a modellare le regioni di associazione allelica.

Diversi studi sono state fatti per determinare il livello medio di LD nel genoma umano (Reich et al. 2001; Gabriel et al. 2002) allo scopo di individuare dei veri e propri blocchi di aplotipi (ovvero successioni di alleli associati sullo stesso cromosoma). Attualmente è in corso un grosso progetto internazionale “The International HapMap Project” (<http://www.hapmap.org/>) il cui scopo è quello di sviluppare una mappa degli aplotipi presenti nel genoma umano (la “HapMap”), che descriva i pattern di variazione della sequenza di DNA in diverse popolazioni (Altshuler et al. 2005). La HapMap servirà a facilitare gli studi di associazione e comprendere meglio i processi evolutivi nell'uomo.

Nelle popolazioni piccole e isolate si riscontrano ampie regioni in LD rispetto alle popolazioni aperte (Kruglyak 1999; Varilo et al. 2000; Ardlie et al. 2002). In realtà, queste popolazioni non presentano alcuna variazione nel tasso di mutazione o di ricombinazione rispetto alle popolazioni aperte, ma alcune ricombinazioni non sono visibili a causa della presenza di regioni identiche per discendenza.

### 1.2.3 - Struttura delle popolazioni isolate

Nelle popolazioni isolate, gli individui che superano il collo di bottiglia che ha generato l'isolamento, inizialmente conservano parte della diversità presente nella popolazione parentale. Dopo un certo numero di generazioni, durante le quali la popolazione si espande attraverso accoppiamenti tra individui della popolazione stessa e in assenza di migrazioni, si verifica un aumento del tasso di inbreeding con

conseguente riduzione della diversità allelica. A questo contribuisce anche la deriva genetica che agisce soprattutto nelle prime generazioni. Le popolazioni isolate sono quindi caratterizzate da un ridotto numero di fondatori, da un aumento del numero di loci in omozigosi, da una riduzione della diversità allelica e da ampie regioni in linkage disequilibrium.

### 1.3 - I tratti complessi

Un tratto complesso è determinato da una componente genetica non strettamente Mendeliana (dominante, recessiva o sex-linked) e da una componente ambientale. Il fenotipo dei tratti complessi è il risultato dell'interazione di due o più geni tra loro e di questi ultimi con fattori ambientali. Esempi di tratti complessi sono le malattie cardiovascolari, l'ipertensione e il diabete (Palmer and Cardon 2005).

Nel caso di malattie monogeniche, un singolo gene è necessario e sufficiente a produrre un fenotipo clinico e l'impatto del gene sul "rischio genetico" di sviluppare la malattia è lo stesso in qualsiasi individuo. Nel caso di malattie complesse, una determinata combinazione di alleli a più geni, può risultare in una "predisposizione genetica" per un certo fenotipo clinico, ma spesso le mutazioni genetiche non sono né sufficienti né necessarie a spiegare il fenotipo: lo stile di vita e l'ambiente sono talvolta i principali responsabili dell'insorgenza di queste malattie.

Nelle malattie complesse ciascun gene agisce in base al background genetico in cui si trova (Peltonen and McKusick 2001). Ad esempio, supponendo che in una data popolazione sia stato identificato un certo numero di geni responsabili di un fenotipo complesso e osservando le singole famiglie, si vedrà che l'impatto di uno stesso gene può essere totalmente differente in ciascuna di esse. Inoltre, un gene raro nella popolazione, può avere un forte impatto nel determinare il fenotipo in una famiglia, ma a causa della bassa frequenza nella popolazione, l'effetto complessivo in essa può essere trascurabile. Al contrario, un gene frequente nella popolazione può avere un effetto completamente trascurabile in una particolare famiglia (**fig6**).

### 1.3.1 - Utilizzo delle popolazioni isolate per lo studio dei tratti complessi

Trascurando la componente ambientale, esistono due modelli estremi che spiegano quale possa essere il background genetico dei tratti complessi. In un primo modello, un elevato numero di alleli piuttosto frequenti nella popolazione può interagire per determinare la predisposizione al tratto; questi alleli mostreranno bassa penetranza, sebbene conferiscano un'alta probabilità di manifestare il tratto. Nel secondo modello, al contrario, interviene un ridotto numero di alleli che manifestano un'alta penetranza e che sono piuttosto rari nella popolazione, ma significativamente frequenti in gruppi ristretti di individui derivanti da un effetto fondatore. In entrambi i modelli il numero di loci coinvolti e la diversità allelica saranno ridotti nelle popolazioni isolate (Varilo and Peltonen 2004). Questo avvicina le modalità di trasmissione a quelle mendeliane classiche, riducendone la complessità. L'effetto fondatore inoltre conferisce all'analisi di associazione un potere elevato, in quanto le varianti geniche introdotte dai fondatori sono assortite nella popolazione attuale che condivide ampie regioni IBD (**fig7**).

Altri fattori che rendono vantaggioso l'uso di queste popolazioni nello studio dei tratti complessi sono: la riduzione del background ambientale per la presenza di uno stile di vita e di un ambiente omogenei, la possibilità di ricostruire un'ampia genealogia ed un'accurata definizione del fenotipo (Peltonen et al. 2000).

Il limite nell'uso di queste popolazioni consiste fondamentalmente nella incertezza di replicare lo studio in altre popolazioni, a causa delle caratteristiche stesse del tratto complesso, il quale si manifesta in ogni background genetico a causa di una componente diversa (Hirschhorn and Daly 2005). Tale limite tuttavia è ripagato dalla possibilità di contribuire all'identificazione delle singole componenti del network che determinano il fenotipo complesso.

## 1.4 La popolazione di Campora

### 1.4.1 - Cenni storici

Campora è un paese che si trova nell'entroterra della regione del Cilento, territorio situato nella zona meridionale della Campania, delimitato a nord dai monti Alburni, ad est dal Vallo di Diano, a sud dal confine con la Basilicata. Le popolazioni cilentane sono dedite prevalentemente all'agricoltura con larga diffusione nelle vallate e nelle zone pianeggianti. Tutto il Cilento è stato luogo di conquiste, nel corso dei secoli, da parte di diversi popoli che si sono avvicinati in tutto il territorio, infatti antiche e numerose sono le testimonianze delle varie popolazioni: Greci, Romani, Bizantini, Longobardi, Arabi, Spagnoli e Francesi. La civiltà contadina del Cilento ha seguito vicende alterne di sviluppo, stasi e progresso in relazione all'influenza dei suoi coloni e dominatori tra cui Angioini, Aragonesi, Borboni. La popolazione cilentana risiede tutt'oggi in numerose cittadine e villaggi distribuiti tra colline e vallate, o dispersi tra i monti, secondo l'eterogenea geografia di questo territorio, in cui si ritrovano piccole comunità con strutture socio-culturali diversificate, in cui risultano dispersi articolati patrimoni culturali e storici dovuti alle stratificazioni di tante civiltà. Questa condizione storico-geografica ha provocato, nel corso dei secoli, un inevitabile isolamento dei centri situati nelle zone interne, aggravato dall'assenza di adeguate vie di comunicazione, aspetto che ha caratterizzato queste zone fino a pochi decenni fa. Campora è, appunto, uno di questi piccoli comuni collinari, che sorge nel cuore del Cilento, nella zona dell'alto Calore. Situato su una collina calcarea, ad una altitudine compresa tra i 380 m. e i 900 m., sul livello del mare, l'insediamento urbano si trova a 528 metri sopra il livello del mare. Il territorio comunale di Campora fa parte del "Parco nazionale del Cilento".

Il nome "Campora" è di derivazione greca: "Canphora". Nel corso del tempo il nome originario si è trasformato in "campo ora", dal latino "campus orationis" (campo, zona di preghiera).

Il primo nucleo abitato del paese si costituì tra la fine del X e gli inizi dell'XI secolo, intorno al monastero di S.Giorgio di Campora di Sicura, un cenobio basiliano. Secondo lo storico P.Ebner è in questo periodo che nella regione cilentana giungono dai Balcani o dall'Oriente monaci italo-greci, nel tentativo di sfuggire all'invasione araba. Questi monaci impiantano, dovunque arrivano, intorno all'abazia un'azienda agricola. Insegnano la coltivazione del gelso, della vite, dell'ulivo alle popolazioni locali. E' dovuta alla civiltà basiliana, nel Cilento, la riattazione di strade, la bonifica di zone allagate, questi monaci costruiscono villaggi agricoli alcuni dei quali destinati a svilupparsi in centri urbani.

Le notizie storiche sul villaggio di Campora sono piuttosto frammentarie e sono dovute, in particolare, agli studi dello storico P.Ebner: la vita di Campora è stata legata soprattutto alla funzione di controllo che questo centro abitato poteva svolgere sull'importante via di transito che da Vallo della Lucania passava da queste terre per poi proseguire per il Vallo di Diano. Le prime notizie certe su Campora partono dal 1131 quando, da un documento, si evince che il re normanno Ruggiero II confermò ai padri basiliani di Grottaferrata il possesso della torre della Grancia di Sant'Arcangelo. Come il resto del territorio la storia di Campora è legata all'avvicinarsi delle diverse colonizzazioni, durante le quali al villaggio come ai numerosi borghi vicini, vengono imposti numerosi e disparati governi secondo la volontà dei regnanti che si sono susseguiti in tutto il Meridione. Nel 1656 Campora fu colpita dalla peste, che ne decimò la popolazione, lasciando verosimilmente solo 140 sopravvissuti. Campora aderì alle rivolte patriottiche e scrisse importanti pagine del Risorgimento: nel febbraio del 1821, a Campora, come nel resto del Cilento, i contadini occuparono le terre padronali e in seguito la popolazione reagì violentemente al regime poliziesco dei Borboni. La popolazione subì un forte calo a cavallo fra Otto e Novecento: molti camporesi, infatti, a causa della crisi agraria furono costretti a emigrare. Secondo la tradizione, i camporesi si dedicano ancora oggi all'agricoltura - in particolare alla coltivazione dell'ulivo - e in misura minore alla pastorizia.

#### 1.4.2 - La genealogia di Campora

La genealogia di Campora a partire dalla seconda metà del XVI secolo ad oggi, è stata pazientemente ricostruita e resa disponibile in formato digitale. I dati della genealogia derivano dai registri anagrafici comunali per gli ultimi secoli e principalmente dagli archivi parrocchiali e diocesani fino al XVI secolo.

Il pedigree include 10737 individui (5256 femmine e 5441 maschi) distribuiti in 15-17 generazioni. La durata media di una generazione, calcolata dal pedigree, è di 32.5 anni.

#### 1.5 Scopo del lavoro di tesi

Il progetto “Parco Genetico del Cilento” è nato circa tre anni fa con lo scopo di mappare geni responsabili di tratti complessi, utilizzando popolazioni umane isolate come sistema modello. Il progetto prevede lo studio di diverse popolazioni residenti nel Parco Nazionale del Cilento e vallo di Diano, che sembrano presentare le caratteristiche tipiche delle popolazioni isolate in base ad evidenze storiche e geografiche.

La prima fase del progetto è consistita nella caratterizzazione della struttura genetica delle popolazioni prescelte, allo scopo di verificare che queste possiedano tutti i requisiti per essere considerate isolate. Si è proceduto inoltre alla raccolta di dati relativi al fenotipo degli individui appartenenti alla popolazione e alla genotipizzazione degli stessi, allo scopo di correlare i dati fenotipici a quelli genotipici.

Nell’ambito del progetto “Parco Genetico del Cilento”, mi sono occupata dello studio della popolazione di Campora, allo scopo di stabilire se la popolazione possiede tutti i requisiti tali da poter essere considerata una popolazione geneticamente isolata.

In particolare ho verificato:

1) se in passato siano accaduti fenomeni che possano aver agito come cause dell’isolamento

2) se la popolazione attuale derivi da un ridotto numero di fondatori

3) se sono presenti nel genoma i segni dell'isolamento quali la riduzione della diversità allelica e la presenza di ampie regioni in LD.

## 2 - RISULTATI

### 2.1 - Notizie di demografia

#### 2.1.1 Andamento della popolazione

La maggior parte delle informazioni storiche sul Cilento, sono state raccolte dallo storico Pietro Ebner. Nei suoi libri sono contenute quasi tutte le notizie storiche riguardanti Campora. Da questi libri derivano i dati, relativi al numero di abitanti di Campora dal 1500 al 1970, mostrati in **fig8**. Alcuni di essi sono stati raccolti utilizzando le informazioni contenute negli “Stati delle Anime”, documenti che venivano compilati dal parroco, in occasione della benedizione pasquale, per la quale egli si recava in tutte le case del paese. In tale occasione veniva censito il numero di “fuochi” ovvero case abitate da famiglie comprendenti più di una generazione e il numero delle “anime”, ovvero abitanti di ciascun fuoco.

Sempre in **fig8** è possibile osservare una drastica riduzione del numero di abitanti in corrispondenza della peste del 1656.

#### 2.1.2 Esogamia

Tra i registri parrocchiali sono presenti i registri di matrimonio, nei quali vengono annotate le notizie relative agli sposi. Dai dati relativi alla loro provenienza si deduce se il matrimonio è endogamico (tra individui dello stesso paese) o esogamico (tra individui di paesi diversi). Dalle informazioni di questi registri ho calcolato il numero di matrimoni esogamici in intervalli di tempo di trenta anni (durata media di una generazione) ed ho calcolato quale percentuale rappresenta rispetto al numero totale

di matrimoni. I risultati sono mostrati in **fig9** in cui si vede come questa percentuale sia rimasta piuttosto bassa (valore massimo pari al 20%) fino alle ultime due generazioni.

## 2.2 - Analisi del pedigree

### 2.2.1 Linee genealogiche

Le linee genealogiche materne (matrilinee) o paterne (patrilinee) consistono nell'insieme di individui che discendono da un unico fondatore (ovvero un "capostipite", un individuo per cui non esistono nella genealogia né madre né padre) per via materna o paterna. Tali linee si ottengono dal pedigree a partire dagli individui che si trovano ad un estremo di esso e seguendo in modo continuo, ascendente o discendente, le relazioni madre-figlia o padre-figlio (**fig10**). Nel nostro caso le linee sono state ricostruite a partire dalla popolazione vivente in modo ascendente.

In **tab2** è mostrata la distribuzione temporale dei capostipiti nel pedigree. L'ultimo secolo include un altissimo numero di capostipiti, evento ascrivibile alla maggiore quantità di informazioni reperibili per questo periodo. Inoltre è stato riscontrato, anche esaminando i dati riguardanti l'esogamia (**fig9**), il verificarsi di un aumento del flusso immigratorio in quest'ultimo secolo, che ha contribuito ad incrementare il numero di individui che hanno dato origine a nuove linee.

### 2.2.2 Definizione della popolazione nativa

Attraverso l'analisi delle linee genealogiche, ho potuto definire quale frazione della popolazione può essere considerata "nativa" di Campora, cioè residente nel villaggio da almeno tre generazioni. Infatti, i dati relativi agli individui appartenenti al pedigree, a causa delle modalità in cui sono stati raccolti, includono anche persone che recentemente sono immigrate in questo villaggio le quali non sono state considerate come native. Dai calcoli effettuati sul pedigree, è emerso che la durata media della generazione, intesa come età media in cui un individuo si riproduce, è di circa 30 anni, sia per i maschi che per le femmine, per cui tre generazioni corrispondono a circa un secolo. Ho quindi definito nativi tutti i discendenti (1184 individui) da capostipiti nati

prima del 1890. Ho quindi calcolato la percentuale di individui nativi rispetto alla popolazione anagrafica vivente ed ho così determinato che il 69% della attuale popolazione anagrafica femminile e il 76% dell'attuale popolazione anagrafica maschile, possono essere considerati nativi di Campora (fig11). Tutte le successive analisi si riferiscono a questa fascia di popolazione. In fig12 è mostrata la ripartizione della popolazione nativa vivente in classi di età.

### 2.2.3 Inbreeding nella popolazione nativa

Il valore del coefficiente di inbreeding (f), moltiplicato per 100, costituisce la probabilità che in un individuo, due alleli ad un locus siano IBD. Ho determinato il valore medio di f della popolazione nativa utilizzando la genealogia. Questa metodologia prevede il calcolo del numero di generazioni che ci sono nel percorso circolare (loop) che collega l'individuo di cui si calcola f all'antenato comune ad entrambi i genitori (fig3). Maggiore è il numero di generazioni contenute nel pedigree, maggiore è la possibilità di avere dei loops completi. Il calcolo è stato fatto, quindi, sull'intero pedigree, che comprende circa 10737 individui, mediante l'applicazione PEDSYS, utilizzando la funzione KINSHIP. Questo software restituisce come risultato un valore di coefficiente di inbreeding per ogni individuo. Tale valore è uguale a zero in assenza di inbreeding.

Tra gli individui nativi viventi il 77% dei maschi e l'87% delle femmine presentano un valore di f diverso da zero. In questi due gruppi la media di f è pari a  $0,0055 \pm 0,0074$  e  $0,0075 \pm 0,0106$  rispettivamente, mentre nel totale della popolazione nativa vivente f medio per entrambe i sessi è pari a  $0,0065 \pm 0,0091$  (**tab3**).

### 2.2.4 Distribuzione degli individui nelle linee genealogiche

I 1184 individui identificati come nativi viventi, si possono raggruppare in 49 linee femminili e 69 linee maschili indipendenti (**tab2**). Negli istogrammi in fig13 sono indicati il numero di discendenti per ciascun capostipite, sia maschile che femminile. In realtà la popolazione attuale non discende da 49 femmine e 69 maschi, in quanto tale dato presenta un limite legato ai dati genealogici dipendente dalla disponibilità e dallo

stato degli archivi, nonché dal fatto che non è possibile reperire nessun tipo di dato anteriore al 1600. L'analisi delle linee genealogiche mi ha permesso di arrivare ad una conclusione parziale: se i dati anagrafici fossero stati disponibili per molti altri secoli addietro allora la conclusione probabilmente sarebbe stata diversa.

### **2.3 - Determinazione del numero di fondatori della popolazione di Campora**

Una delle caratteristiche delle popolazioni isolate è che queste derivano da un ridotto numero di fondatori. E'ragionevole pensare che alcune delle linee genealogiche convergano in un unico fondatore e quindi il numero reale dei fondatori sia inferiore a quello delle linee genealogiche. Questa ipotesi si è rivelata corretta: la determinazione in-silico di tali linee è stato il punto di partenza per effettuare l'analisi molecolare, utilizzando rispettivamente alcuni marcatori del DNA mitocondriale e del cromosoma Y. Attraverso questa analisi abbiamo potuto definire con maggiore accuratezza il numero di fondatori da cui è derivata la popolazione attuale di Campora. La scelta di utilizzare marcatori mitocondriali e del cromosoma Y è ovviamente giustificata dalle peculiari modalità di trasmissione di queste due parti del corredo genomico umano.

#### **2.3.1 Determinazione del numero di linee fondatrici femminili**

Tradizionalmente il DNA mitocondriale è stato caratterizzato mediante polimorfismi di restrizione (Cann et al. 1982; Cann and Wilson 1983). La combinazione di determinati polimorfismi definisce un aplogruppo, il quale può essere usato per caratterizzare una popolazione. In effetti quasi tutti i genomi mitocondriali umani appartenenti ai diversi gruppi etnici sono stati caratterizzati più o meno in dettaglio. Esiste una classificazione degli aplogruppi (indicati con una lettera maiuscola in fig14), e la loro distribuzione rispetto alle aree geografiche e alle popolazioni è nota, come mostrato in fig15 (Mitomap, [www.mitomap.org](http://www.mitomap.org) e mtDB [www.genpat.uu.se/mtDB](http://www.genpat.uu.se/mtDB)).

La popolazione di Campora comprende 576 femmine native viventi ripartite in 49 linee genealogiche indipendenti. I campioni di DNA raccolti coprono 46 linee

genealogiche che rappresentano il 98 % delle femmine native viventi. La strategia per determinare il numero delle “linee fondatrici femminili” consiste nel verificare quanti diversi aplogruppi mitocondriali sono presenti nella popolazione (Angius et al. 2001). Si parla di linee femminili fondatrici piuttosto che di femmine fondatrici in quanto esiste anche la possibilità che due femmine fondatrici diverse abbiano avuto lo stesso aplogruppo.

Ho caratterizzato la sub-popolazione femminile mediante analisi di restrizione allo scopo di verificare la presenza dei principali aplogruppi europei ovvero H, T, U, V, W, X, I, J, K, M (Torroni et al. 1996). Ho scelto di analizzare due o più femmine per ciascuna linea genealogica; il campione così assemblato risulta rappresentativo di tutta la sub-popolazione femminile.

Mediante il confronto tra campioni appartenenti alla stessa linea, ho accertato la consistenza di ciascuna linea genealogica, come conferma e controllo del calcolo effettuato in-silico. Confrontando campioni di linee diverse, ho invece determinato la presenza di 6 su 10 degli aplogruppi analizzati (non sono stati trovati individui appartenenti agli aplogruppi V, W, I). Tuttavia 10 linee, che corrispondono al 21% della popolazione femminile nativa vivente, non ricadono in nessuno degli aplogruppi analizzati e sono stati classificati come “Others”. I risultati sono schematizzati in fig16a, dove per ogni aplogruppo viene indicata la corrispondente percentuale di femmine native. Come atteso, la maggior parte degli individui (62%) appartiene all’aplogruppo H che è l’aplogruppo maggiormente presente in Europa. Ciò non implica necessariamente che il 62% delle femmine native derivi da un'unica linea fondatrice. Ciascun aplogruppo può essere suddiviso in sotto-aplogruppi. Allo scopo di definire i sotto-aplogruppi e anche per le linee che non è stato possibile classificare mediante l’analisi di restrizione, ho analizzato ulteriori polimorfismi di sequenza presenti in una particolare regione del genoma mitocondriale: la regione ipervariabile I (HVR-1). Tale regione è così denominata perché soggetta ad un tasso di mutazione più alto rispetto al resto del genoma mitocondriale a causa della vicinanza di questo tratto di genoma con una fonte di radicali liberi presente nel mitocondrio. La HVR-1 è stata caratterizzata in diverse popolazioni ed è in fase di allestimento un database che contiene e classifica i polimorfismi di sequenza di questa regione (HVR Database [www.hvrbase.org/](http://www.hvrbase.org/)).

Ho sequenziato la HVR-1 degli stessi individui che avevo analizzato per restrizione e ancora ho confermato la consistenza di sequenza tra gli individui di una stessa linea e identificato 48 siti polimorfici che mi hanno permesso di rifinire l'analisi di restrizione (**tab4**). Complessivamente ho identificato 29 differenti sequenze che corrispondono ad altrettante linee femminili fondatrici. In fig16b, è indicata la percentuale di femmine native viventi appartenente a ciascuna delle 29 linee fondatrici. Il numero accanto ad ogni lettera indica il sotto-aplogruppo. Dodici linee su ventinove hanno dato ciascuna origine ad una discendenza che comprende un numero di individui pari a meno dell'1% della popolazione femminile nativa vivente e queste sono state raggruppate nella categoria "<1%". Questa categoria corrisponde al 5% del gruppo considerato. Non è stato possibile caratterizzare il 2% delle femmine, appartenenti a linee genealogiche non campionate. Escludendo queste due fette, pari al 7%, si può affermare che il 93% della popolazione femminile nativa vivente di Campora deriva da 17 linee femminili fondatrici.

### 2.3.2 Determinazione del numero di linee fondatrici maschili

Per determinare il numero di linee fondatrici maschili ho seguito una strategia simile a quella utilizzata per la determinazione delle linee fondatrici femminili analizzando dei marcatori molecolari presenti sul cromosoma Y (Angius et al. 2001). In particolare, ho utilizzato 7 microsatelliti (DYS19, DYS385, DYS389, DYS390, DYS391, DYS392, DYS393) che si trovano sul braccio lungo del cromosoma Y. Questi marcatori costituiscono il set minimo utilizzato per la classificazione degli alplotipi del cromosoma Y (<http://www.ystr.org/>) ed anche per scopi forensi. Ciascun microsatellite possiede un certo numero di alleli, come indicato in **tab5**, ciascuno caratterizzato da un diverso numero di ripetizioni di cui è costituito il microsatellite. La combinazione degli alleli dei diversi loci costituisce l'aplotipo. Esiste una probabilità trascurabile che due individui possiedano lo stesso aplotipo, a causa dell'altissimo numero di combinazioni alleliche possibili per i sette loci, a meno che i due individui non siano discendenti dello stesso fondatore.

I 608 maschi nativi viventi sono raggruppati in 69 linee genealogiche separate. E' stato possibile reperire campioni di individui provenienti da sole 45 linee, alle quali,

tuttavia, appartiene il 90% della popolazione maschile nativa vivente. Il campione di maschi è stato assemblato con le stesse modalità di quello delle femmine per le stesse ragioni, scegliendo due o più individui da ciascuna linea genealogica. Per questo campione ho determinato il genotipo ai loci sopra indicati e i risultati mi hanno permesso di individuare 24 diversi aplotipi (**tab6**) verosimilmente corrispondenti ad altrettante linee fondatrici maschili. Anche in questo caso 4 linee sono presenti con un numero di discendenti minore dell'1% della sub-popolazione maschile nativa totale e sono state pertanto raggruppate nella categoria “<1%” (pari al 2%). Nella categoria ND sono stati inclusi il 10% degli individui che non è stato possibile analizzare. Alla luce di tutto questo è possibile affermare che l'89% della popolazione maschile nativa vivente di Campora deriva da 20 linee maschili fondatrici (**fig17**).

## 2.4 Haplotype Diversity sul cromosoma Y

La Haplotype Diversity (HD) misura la probabilità che due individui presi a caso dallo stesso campione possiedano lo stesso aplogruppo. Quando nel campione considerato tutti gli aplogruppi sono differenti allora il valore numerico della HD è pari ad 1 (popolazione aperta) mentre diminuisce nelle popolazioni inbred. Il valore della HD risente della dimensione del campione e pertanto viene normalizzato per questo parametro.

Ho calcolato il valore della HD utilizzando i dati relativi ai marcatori del cromosoma Y considerati per la determinazione del numero delle linee fondatrici maschili. Questi dati riguardano 98 maschi scelti secondo le modalità riportate nei paragrafi 2.3.1 e 2.3.2. Ho creato 77 gruppi costituiti ognuno da 25-50 persone scelte a caso tra i 98. Per ogni gruppo ho calcolato il valore della HD utilizzando il programma Arlequin (<http://anthropologie.unige.ch/arlequin>). Ho quindi fatto una media delle HD dei 77 gruppi ed ho ottenuto un valore medio pari a  $0.94 \pm 0.02$ . Lo stesso parametro nella popolazione della regione Campania a cui Campora appartiene, è pari a 0.99 (Calcagno et al. 2005). Come controllo ho calcolato la HD considerando un campione di 29 maschi ciascuno appartenente ad una diversa linea fondatrice ed ho trovato un valore di HD ovviamente pari ad  $1.000 \pm 0.0113$ .

## 2.5 Analisi del PATTERN di Linkage Disequilibrium

L'analisi del LD consiste nel valutare l'associazione tra due alleli a due loci differenti. Ho scelto di caratterizzare una regione genomica che è già stata caratterizzata in altre popolazioni, sia isolate che aperte, per poter effettuare un confronto tra queste e i dati di Campora. Per calcolare i parametri che misurano il LD è indispensabile conoscere la fase degli alleli sul cromosoma ovvero quali alleli a due o più loci successivi si trovano sullo stesso cromosoma. La fase allelica si determina conoscendo sia il genotipo dei genitori che dei figli per i loci di-allelici, mentre è già determinata per i loci mono-allelici, come ad esempio le regioni di emizigosi (cromosomi X e Y nei maschi).

### 2.5.1 LD nelle regioni Xq13.3-Xq21.1

Per lo studio del LD sul cromosoma X, ho scelto un campione di 58 individui maschi, nei quali la fase dei marcatori sul cromosoma X è inequivocabilmente determinata. Per questi individui è stato determinato il genotipo dello stesso set di microsatelliti utilizzato in altri studi (Laan and Paabo 1997; Taillon-Miller et al. 2000; Zavattari et al. 2000) che si trovano nelle regioni Xq13.3 e Xq21.1. Alcune caratteristiche di questi marcatori sono descritte in **tab7**. Ho quindi calcolato il grado di associazione allelica tra marcatori per tutte le coppie possibili di alleli mediante un test di Fisher implementato nell'algoritmo del modulo Haploxt del software GOLD. Questo test restituisce un p-value, associato all'ipotesi zero di associazione allelica, che corrisponde alla probabilità che tale associazione si verifichi a caso. Pertanto risultano significativamente in LD le coppie di marcatori che presentano un p-value molto basso, ovvero una bassa probabilità di essere associate per caso. Il massimo p-value, oltre il quale l'associazione tra gli alleli non è significativa, è pari a 0.05.

I risultati relativi al p-value sono mostrati nella **tab8**, che contiene anche dati già pubblicati relativi sia a popolazioni considerate isolate (Gavoi, Saami) che a popolazioni considerate aperte.

### 2.5.2 Linkage Disequilibrium nel genoma

Avendo a disposizione il genotipo a 1060 loci, distribuiti in tutto il genoma, per tutta la popolazione, ho effettuato un'analisi di LD nell'intero genoma, in un campione di circa 400 cromosomi. La dimensione del campione è stata scelta in base alle recenti indicazioni sullo studio del LD in popolazioni isolate (Tenesa et al. 2004), in modo tale da assicurare l'assenza di bias dovuti proprio alla dimensione e alla tipologia di popolazione. Per ciascun cromosoma ho calcolato i valori di LD tra tutte le possibili coppie di marcatori. Ho poi suddiviso le coppie di marcatori in base alla distanza che li separa, in classi di 0.5 cM. In ciascuna classe ho calcolato il percento di loci in significativo LD ( $p\text{-value} < 10E-6$ ) ed ho costruito un grafico, mostrato in **fig18**.

## 3- MATERIALI E METODI

### 3.1 - Database

Per Campora è stato generato un database relazionale in PostgreSQL che contiene tutti i dati relativi alla genealogia. Il database è composto di tabelle che possono essere suddivise in due categorie: tabelle che descrivono le informazioni essenziali in relazione alla popolazione (ad esempio pedigree) e tabelle che descrivono le misure fenotipiche effettuate (analisi di laboratorio, risultati delle visite mediche, etc.) e le informazioni di genotipizzazione. In merito alla genotipizzazione, ad oggi, per tutti gli individui di Campora è stato determinato il genotipo ad oltre mille loci distribuiti nel genoma, come indicato in **tab9**

Ho estrapolato tutti i dati relativi alle notizie storiche dal database di Campora, interrogandolo con delle query di Acces (Microsoft)

### 3.2 - Ricostruzione delle linee genealogiche

Nel laboratorio in cui lavoro è stato scritto un algoritmo in Perl che calcola le linee genealogiche partendo da ogni individuo vivente e procedendo in modo da includere nello stesso gruppo le femmine (maschi) ad esso direttamente collegate se è femmina (maschio), fino a giungere agli individui che senza ulteriori ancestor che sono quindi considerati i capostitipi così come descritto in **fig10**. Ogni gruppo rappresenta una linea genealogica e non vi è sovrapposizione di individui tra le linee.

### 3.3 - Campionamento e preparazione di DNA

Il campionamento del DNA è stato organizzato nel villaggio di Campora nel corso di diversi mesi. Il DNA degli individui analizzati è stato estratto a partire da 10 ml di sangue periferico secondo la metodologia del salting-out utilizzando i kit per l'estrazione del DNA Flexigen (QIAGEN). Tutti i DNA sono stati catalogati elettronicamente e conservati a 4°C.

### 3.4 - Files formato linkage

La maggior parte dei programmi che ho utilizzato richiede come input dei files in formato linkage. Questi sono files di testo in cui ogni linea corrisponde ad un individuo. Per ciascun individuo sono specificati in ordine ed in colonne successive: l'identificativo del pedigree di appartenenza, il proprio identificativo (ID), l'identificativo del padre (IDfa), l'identificativo della madre (IDmo), il sesso, lo stato. Questo ultimo parametro, in generale, indica gli individui che il programma deve considerare come probandi. A queste cinque colonne, sempre presenti, vanno aggiunte altre colonne che possono indicare diverse tipologie di dati. Ad esempio dati genotipici: locusA\_allele1, locusA\_allele2, locusB\_allele1, locusB\_allele2, etc. Oppure dati aploipici: locusA\_allele1, locusB\_allele1, etc.

Un esempio di file formato linkage è in **fig19**. Ho preparato tutti i files in formato linkage necessari, a partire dai dati contenuti nel database di Campora.

### 3.5 - Coefficiente di inbreeding

Ho calcolato il coefficiente di inbreeding utilizzando la funzione KINSHIP contenuta nel software PEDSYS ([www.sfbr.org/pedsys/pedsys.html](http://www.sfbr.org/pedsys/pedsys.html)). L'algoritmo di kinship si basa sulla formulazione di Stevens-Boyce (Cothran et al. 1984) secondo cui:

$$f = \sum_{i=1}^n \left( \frac{1}{2} \right)^{n_i} (1 + f_a)$$

dove  $f$  è il coefficiente di inbreeding di un individuo,  $l$  è il numero di loops che si vengono a formare seguendo il percorso da e per l'individuo passando per l'ancestor in comune ai genitori (fig3),  $n_i$  è il numero di individui inclusi nel loop (escluso quello di cui si sta calcolando  $f$ ),  $f_a$  è il coefficiente di inbreeding dell'individuo all'apice del loop (ovvero l'ancestor).

E' necessario fornire al programma in input un file pedigree in formato linkage, il calcolo viene fatto su tutti gli individui del pedigree. Come output si ottiene un file in cui in ogni riga sono riportati i dati relativi ad un individuo e con 4 colonne contenenti ID individuo, ID padre, ID madre e il valore del coefficiente di inbreeding.

### 3.6 - Linkage Disequilibrium nel genoma

#### 3.6.1 - Messa in fase dei marcatori

Il calcolo del linkage disequilibrium prevede che i dati siano sotto forma "aplotipica", ovvero che la successione esatta degli alleli ai differenti loci considerati sia specificata. E' possibile mettere i dati della genotipizzazione in questo formato conoscendo il genotipo dei genitori. Ho utilizzato il software SHOWHAPLO (F.Dudbridge; disponibile su <ftp.hgmp.mrc.ac.uk/pub/linkage>). Il programma accetta un file pedigree in formato linkage. I dati devono essere sotto forma di famiglie nucleari con qualsiasi numero di figli. In output si ottiene un file in cui sono indicati, per ogni figlio, l'aplotipo trasmesso per via materna e quello paterno.

Ho scelto un campione di 200 individui ed ho determinato per ciascun individuo l'aplotipo per ciascun cromosoma a partire dai dati di genotipizzazione contenuti nel database. In **tab9** sono riportati, per ciascun cromosoma, il numero di marcatori utilizzati e la distanza media che li separa.

### 3.6.2 - Calcolo del LD per ciascun cromosoma

Una volta determinato l'aplotipo è possibile procedere al calcolo del LD tra marcatori. A tale scopo ho utilizzato il modulo Haploxt del software GOLD (<http://www.sph.umich.edu/csg/abecasis/GOLD/>) il cui algoritmo si basa sul Test di Fisher. Il software accetta un file di input in cui ogni linea corrisponde ad un individuo e le colonne contengono in ordine l'identificativo dell'individuo e a seguire il genotipo "in fase" dei marcatori che si vogliono analizzare. Il software considera tutte le possibili coppie di marcatori e per ciascuna restituisce un p-value.

### 3.6.3 - Correzione per test multipli

Il p-value va interpretato secondo la soglia che si ritiene opportuna. Normalmente questa soglia ( $\alpha$ ) è lo 0.05 che equivale a dire che esiste il 5% di probabilità che quello che si osserva è casuale e non significativo (errore di tipo I). Nel caso di test multipli tuttavia, è necessario applicare una correzione. Infatti la soglia reale che si osserva è data dalla formula:

$$\alpha \text{ reale} = 1 - (1 - \alpha \text{ che si sta settando})^J$$

dove  $J = n^\circ$  di test che sono stati effettuati. Ad esempio, se si sta settando  $10E-4$  per circa 1000 test, esiste in realtà il 10% di possibilità di commettere un errore di tipo I; se si sta settando  $10E-7$  per lo stesso numero di test questa possibilità si riduce all'1%. Ogni cromosoma va trattato come una serie indipendente di test ed in **tab9** è mostrato il valore della soglia in base al numero di marcatori per ogni cromosoma. Considerando questi valori e dovendo trattare tutti i dati assieme, ho scelto di utilizzare una soglia unica,  $10E-6$ , così come specificato in fig18.

### 3.6.4 - Linkage Disequilibrium sul cromosoma X

Per il calcolo del LD sul cromosoma X ho scelto di utilizzare maschi in modo da aggirare il passaggio della messa in fase dei marcatori. Ho determinato il genotipo a 6 loci (DXS5983, DXS8092, DXS8037, DXS1225, DXS8082, DXS986) nella regione Xq13.3-Xq21.1 in un campione di 58 maschi non relazionati. Ho ottenuto la sequenza

dei primers da Genome Database (<http://gdbwww.gdb.org/>). Le condizioni di amplificazione per PCR sono 94°C per 2', quindi 94°C per 30", da 60°C a 65°C (-0,5°C/ciclo) per 30", quindi 72°C per 30", per 15 cicli, 96°C per 15", 65°C per 30", 72°C per 30", per 20 cicli ed infine 72°C per 4'. I prodotti di PCR sono stati analizzati per elettroforesi capillare sulla macchina MegaBACE1000Flexi (Amersham) e ho determinato i genotipi utilizzando il software Fragment Profiler (Amersham).

Ho calcolato quindi i valori di LD, utilizzando Haploxt come descritto sopra.

### 3.7 - Analisi del DNA mitocondriale

Per l'analisi del DNA mitocondriale ho scelto un campione di 80 femmine secondo le linee genealogiche prendendo almeno due campioni per ogni linea quando è stato possibile. Ho amplificato per PCR 8 regioni del genoma mitocondriale (**tab10**). Il programma di PCR prevede 94°C per 2', quindi 94°C per 30", da 60°C a 65°C (-0,5°C/ciclo) per 30", quindi 72°C per 30", per 15 cicli, 96°C per 15", 65°C per 30", 72°C per 30", per 20 cicli ed infine 72°C per 4'. Ho poi effettuato delle analisi di restrizione sui frammenti di PCR ottenuti, così come indicato in **tab10** per analizzare 10 polimorfismi, la combinazione dei quali definisce i 10 aplogruppi analizzati. Ho controllato la posizione di mappa dei siti utilizzando il database Mitomap ([www.Mitomap.org](http://www.Mitomap.org)). Per ogni digestione di restrizione, 6-8µl di prodotto di PCR sono stati incubati con 5 di enzima, in un volume finale di 20µl, e 8-10µl di prodotto di digestione sono stati visualizzati su gel di agarosio al 2% mediante colorazione con bromuro di etidio e illuminazione con raggi UV.

Ho analizzato i polimorfismi della regione ipervariabile I (HVR-1) nella regione compresa tra i nucleotidi 15940-16383. Ho amplificato per PCR questa regione secondo questo programma: 94°C per 2', quindi 94°C per 30", da 58°C a 63°C (-0,5°C/ciclo) per 30", quindi 72°C per 30", per 15 cicli, 96°C per 15", 65°C per 30", 72°C per 30", per 20 cicli ed infine 72°C per 4'. La reazione di sequenza è stata effettuata su una macchina BigDye Terminator Cycle Sequencing Ready Reaction (Applied Biosystem, Warrington, UK) ed il sequenziamento per elettroforesi capillare è avvenuto sulla

macchina ABI PRISM 377 DNA analyzer (PE Biosystems). Ho esaminato le sequenze utilizzando il software AutoAssembler (Applied Biosystem, Warrington, UK). Ho confrontato la sequenze con quella di riferimento (Anderson et al. 1981; Ingman et al. 2000)

### **3.8 - Analisi del cromosoma Y**

Per l'analisi del cromosoma Y ho scelto un campione di 87 maschi seguendo lo stesso criterio adottato per lo studio del DNA mitocondriale, ma questa volta utilizzando le linee genealogiche maschili.

Nel campione scelto, ho determinato il genotipo a 7 loci (DYS19, DYS385, DYS389, DYS390, DYS391, DYS392, DYS393). Ho prodotto per PCR dei frammenti utilizzando i primers universali pubblicati su Genome Database (<http://gdbwww.gdb.org/>). Le condizioni di PCR sono 95°C per 10', 95°C per 30", 55°C (per i microsatelliti DYS19, DYS390, DYS392, DYS393), 57°C (per i microsatelliti DYS391) and 62°C (per i microsatelliti DYS385, DYS389) per 30", 72°C for 30", per 30 cicli, ed in fine 72°C per 7'. I prodotti di PCR sono stati analizzati sulla macchina MegaBACE1000Flexi (Amersham) e ho determinato i genotipi utilizzando il software Fragment Profiler (Amersham).

Ho calcolato la haplotype diversity (H) e la relative varianza utilizzando il software Arlequin ver2.00 (<http://anthropologie.unige.ch/arlequin>) secondo la formula:

$$H = n / (n-1) (1 - \sum p_i^2)$$

dove n è il numero di campioni e  $p_i$  è la frequenza di ciascun aplotipo.

## **4 - DISCUSSIONE**

### **4.1 - Analisi in Campora delle componenti che caratterizzano le popolazioni isolate**

Durante il lavoro di tesi mi sono occupata dello studio della popolazione di Campora, allo scopo di stabilire se essa può essere considerata un isolato genetico. Le popolazioni isolate sono state utilizzate con successo nello studio delle malattie mendeliane classiche ed attualmente si sono rivelate un valido modello per lo studio dei tratti complessi (Peltonen et al. 1999; Gianfrancesco et al. 2003). Lo studio della popolazione di Campora rientra in un progetto che intende utilizzare le popolazioni isolate per lo studio dei tratti complessi. Ho analizzato in Campora, le principali componenti che caratterizzano le popolazioni isolate.

#### **4.1.1 - Il bottleneck causato dalla peste**

Dal punto di vista genetico, una popolazione si definisce come un gruppo di individui che deriva da uno stesso insieme di antenati. Questo insieme si può definire arbitrariamente o secondo dei criteri, quali ad esempio “tutti gli ancestor che abitavano a Campora nel 1600” o “tutti gli ancestor che sono vissuti nel periodo della peste”. Correntemente si identifica una popolazione in base alla propria collocazione geografica ma, al giorno d’oggi, non è sempre possibile seguire questo criterio. I fenomeni migratori infatti, si sono notevolmente intensificati, ad esempio i registri anagrafici di Campora includono individui non strettamente “originari”, mentre mancano i dati di persone originarie che sono emigrate. A Campora una buona parte della popolazione è emigrata all’estero a partire dagli anni ‘50, nell’immediato dopoguerra, tornando poi spesso a vivere nel proprio villaggio con un consorte straniero. Inoltre, una quantità considerevole di individui provenienti da altre parti dell’Est Europa è entrata a far parte

del villaggio, riflesso di un più generale fenomeno migratorio che riguarda tutta l'Italia. Tutto ciò è evidente dal grafico dell'esogamia (fig9), anche se, molti dei matrimoni esogamici, riguardano individui dei villaggi vicini. Per tutte queste ragioni il primo passo nello studio della popolazione di Campora è stato quello di definire accuratamente la popolazione "nativa" a partire dalla genealogia, differenziandola dalla popolazione anagrafica. Tenendo conto del calcolo che stabilisce che la durata media di una generazione è di 30 anni e che la percentuale di esogamia è rimasta basso fino alle ultime tre generazioni, ho deciso di considerare nativi tutti gli individui che discendono da un antenato facente parte della popolazione di Campora da almeno tre generazioni.

Il villaggio di Campora è stato fondato all'inizio dell'XI secolo dai monaci greci in fuga dalle invasioni arabe ed il primo nucleo abitato era costituito dalla gente locale impegnata nelle attività agricole del monastero. Nel XVII secolo l'epidemia di peste, che stava devastando l'Italia, raggiunse anche questa popolazione che venne duramente colpita. Secondo fonti storiche (Ebner), solo 140 persone sopravvissero, come mostrato in fig8. Questo evento rappresenta un collo di bottiglia, al quale è seguito un periodo di espansione della popolazione in condizioni di isolamento. Non ci sono notizie di immigrazioni avvenute per ripopolare il villaggio, evento che si è verificato in altri villaggi decimati dalla peste, come Montaguto nella Valle Ufita o Carlantino in Puglia. La crescita della popolazione a partire dai sopravvissuti alla peste, è avvenuta in condizioni di isolamento. In passato Campora è stata isolata principalmente per ragioni di tipo geografico: non era sito di transito per raggiungere altri villaggi, né sede di fiere o mercati. A tutto ciò va aggiunta la tendenza degli individui a spostarsi poco dal villaggio nativo.

Ho accertato che la popolazione nativa vivente derivi effettivamente dal bottleneck calcolando quale frazione di essa provenga da antenati che sono nati prima o durante il periodo della peste. Per questo calcolo mi sono servita delle linee genealogiche ed ho trovato che il 74% dei maschi e l'82% delle femmine native viventi derivano da questo gruppo di antenati (fig20). Esiste l'eventualità che questa percentuale sia maggiore in quanto tutti gli antenati nati dopo il periodo della peste potrebbero a loro volta avere altri antenati contemporanei alla peste ma dei quali non abbiamo alcuna traccia nella genealogia.

#### 4.1.2 - Le linee fondatrici

La genealogia si è rivelata uno strumento indispensabile per lo studio corretto della popolazione in quanto tutti i campioni di studio sono stati assemblati tenendone conto. Infatti, date le dimensioni della popolazione e la sua peculiarità di essere inbred, non è possibile fare un campionamento casuale e allo stesso tempo rappresentativo, senza correre il rischio di includere persone imparentate a scapito di altre. Ad esempio nel calcolo delle linee fondatrici sarebbe stato necessario analizzare un campione di dimensioni doppie per raggiungere lo stesso risultato ed anche in questo caso non ci sarebbe stata l'assoluta certezza di includere tutti gli alpotipi presenti. Per determinare quanti fondatori hanno dato origine alla popolazione attuale, ho utilizzando un campione di studio composto da almeno due individui presi da ciascuna linea genealogica. Questo campione è stato analizzato seguendo una metodologia descritta in letteratura.

Lo studio del DNA mitocondriale mi ha permesso di risalire al numero di linee fondatrici femminili. Nell'analisi sono stati inclusi i principali aplogruppi europei e l'analisi è stata rifinita con una serie di polimorfismi della regione ipervariabile. I dati combinati dell'analisi di restrizione e di sequenza mi hanno permesso di stabilire che il 93% delle femmine native viventi deriva da 17 linee fondatrici. È interessante osservare che alcune linee fondatrici non appartengono a nessuno degli aplogruppi europei. Questo suggerisce la presenza di fondatrici provenienti da regioni extra-europee e sarebbe interessante identificare qual'è la provenienza di queste linee attraverso lo studio ad esempio della sequenza nella regione ipervariabile. Lo studio dei polimorfismi del cromosoma Y mi ha permesso, invece, di raggruppare il 90% dei maschi nativi viventi in 20 linee fondatrici diverse.

In definitiva il numero di fondatori della popolazione attuale (1184 individui) è al di sotto dei 50. Una caratteristica importante delle popolazioni isolate è che derivano da un ridotto numero di fondatori. Il range entro cui deve variare questo numero, per una popolazione delle dimensioni di Campora, è compreso tra 10 e 100 (Wright et al. 1999). Posso quindi affermare che, da questo punto di vista, Campora rientra nella tipologia delle popolazioni isolate

#### 4.1.3 - L'effetto della deriva genetica

L'effetto della deriva genetica è proporzionale alle dimensioni della popolazione; è facile quindi immaginare quale enorme effetto essa possa esercitare sulle popolazioni isolate, a causa del ridotto numero di individui che ne fanno parte. La deriva genetica è una grandezza difficile da misurare (Helgason et al. 2003). Non è stata ancora elaborata una vera e propria metodologia per effettuare una misura diretta, tuttavia è possibile valutarne l'azione attraverso l'identificazione delle cause e il riscontro dei suoi effetti. Osservando i grafici che mostrano il contributo di ciascun antenato alla popolazione vivente, in termini di numero di discendenti (fig13), si vede come il range, entro cui questo varia, sia ampio. Questa indicazione è confermata dal calcolo della deviazione standard, il cui valore supera di gran lunga quello della media (**tab11**). Questa oscillazione del tutto casuale nel numero di discendenti e le dimensioni ridotte della popolazione possono essere considerati, per la popolazione di Campora, l'evidenza delle circostanze che causano deriva genetica. In definitiva, i genomi di antenati con molti discendenti, sono presenti nella popolazione attuale un numero di volte significativamente più alto di altri, ovvero, alcune varianti alleliche sono significativamente più frequenti di altre. E' anche possibile valutare l'effetto della deriva genetica osservando il valore di haplotype diversity calcolato sul cromosoma Y (**tab12**). Un valore basso di haplotype diversity è indicativo di una ridotta variabilità allelica: alcuni alleli sono stati persi mentre altri si sono fissati. Ho confrontato il valore di HD calcolato per Campora con i valori relativi agli stessi marcatori, calcolati per altre popolazioni sia isolate che aperte (Kittles et al. 1998; Thomas et al. 1998; Helgason et al. 2000). Osservando i dati in **tab12** si evince chiaramente che Campora presenta valori di haplotype diversity, in questa regione genomica, paragonabili a quelli di altre popolazioni isolate (Finish, Ashkenazic Jewish).

#### 4.1.4 - La popolazione di Campora è una popolazione inbred

Gli isolati genetici differiscono tra loro sotto alcuni aspetti, uno di questi è il grado di isolamento. Per alcune popolazioni la condizione di isolamento è estrema, ad esempio due popolazioni che si sono separate dalla chiesa cattolica nel XVI secolo, gli Amish e gli Hutterites. I primi vivono principalmente in Pennsylvania e sono originari

della Germania; ancora oggi vivono completamente separati dal resto del mondo e da qualsiasi forma di tecnologia. Gli Hutterites si sono stabiliti principalmente in Canada e, come gli Amish, hanno adottato uno stile di vita ispirato alla loro religione che si basa su principi quali l'umiltà, il lavoro, la famiglia e la comunità. L'isolamento di queste due popolazioni è molto spinto ed è di origine religiosa. Un altro esempio di isolamento estremo è quello che si è verificato in passato in India in seguito alla suddivisione della popolazione in caste. La separazione in sub-popolazioni, supportata da dati genetici (Bamshad et al. 2001), si può paragonare ad una condizione di isolamento dovuta a fattori sociali. L'isolamento può essere anche meno spinto, come nel caso della popolazione dell'Islanda, in cui l'effetto dell'isolamento geografico, è parzialmente compensato dalle dimensioni della popolazione. Il risultato è una forma di isolamento complessivamente "meno spinto". Il parametro che meglio riflette questo differente grado di isolamento è il coefficiente di inbreeding. Esso tiene conto della genealogia e dunque di tutto ciò che è accaduto nel corso della crescita della popolazione. Nella popolazione nativa di Campora, il valore medio del coefficiente di inbreeding è  $0.007 \pm 0.0091$ . Tale valore, è particolarmente significativo se comparato a quello di altre popolazioni in cui l'isolamento è estremo, quali gli Hutterites, dove  $f = 0.034 \pm 0.015$  (Weiss et al. 2005).

Campora si colloca dal punto di vista dell'inbreeding, in una situazione intermedia, così come atteso. Sebbene si siano verificate tutte le circostanze che portano all'isolamento, nessuna di queste è stata così estrema, tuttavia, il valore riscontrato indica una differenza sostanziale rispetto ad una popolazione aperta.

#### 4.1.5 - Ampiezza delle regioni in Linkage Disequilibrium

Le popolazioni isolate mostrano più ampie regioni in Linkage Disequilibrium a causa della presenza di ampie regioni IBD. In queste zone, infatti, non è possibile accertare eventi di ricombinazione.

Ho verificato il comportamento della popolazione di Campora in una regione del cromosoma X considerata come una delle più interessanti dal punto di vista della storia demografica di una popolazione (Laan et al. 2005). Come mostra la tab9 nel campione

di Campora i valori di LD sono significativi anche tra marcatori separati da grandi distanze, così come nelle altre popolazioni isolate.

La regione del cromosoma X che ho analizzato è una regione in cui la frequenza di ricombinazione è piuttosto bassa, come mostra il confronto tra valori di distanze genetiche e fisiche mostrato in tab8. Questa regione pertanto è “di per sé una regione che tende ad essere in LD. Ho voluto quindi estendere l’analisi a tutto il genoma, includendo in questo modo anche regioni con una frequenza di ricombinazione più alta. Dall’insieme dei calcoli di LD effettuati per ciascun cromosoma ho ricavato un pool di dati che ho ordinato in base alla distanza tra le coppie di marcatori. Osservando il grafico in fig18 è chiaro come il LD decresce con la distanza tra marcatori così come atteso. Inoltre, di tutte le 25413 possibili coppie di marcatori, 410 sono in significativo LD ed il 75% di queste sono separate da una distanza fino a 6cM. Il risultato più interessante lo si ottiene comparando questa distribuzione con quella relativa ad un campione del CEPH (Centre d'Etude du Polymorphisme Humain, <http://www.cephb.fr/>) che rappresenta un campione estratto da una popolazione aperta (Huttley et al. 1999). Nel grafico del campione CEPH mostrato in fig21 nel primo bin il percento di loci in rilevante LD è pari 18, mentre lo stesso valore nel campione di Campora è più che raddoppiato. Inoltre nel grafico CEPH si raggiunge un plateau del 6% a partire da marcatori distanti 1,5 cM. Lo stesso valore, il 6%, nel campione di Campora si raggiunge a partire da 6 cM.

Tutti questi dati indicano chiaramente che, nel genoma della popolazione di Campora l’ampiezza delle regioni in LD è mediamente più alta rispetto alle popolazioni aperte anche nelle regioni con bassa frequenza di ricombinazione.

#### **4.2 - Considerazioni conclusive sulla popolazione di Campora**

Nella popolazione di Campora, l’analisi delle principali componenti che contraddistinguono le popolazioni isolate, indica che questa popolazione possiede le caratteristiche peculiari delle popolazioni isolate. L’entità degli eventi che ne hanno

modellato la crescita, ha determinato in questa popolazione un livello di isolamento che si è ridotto a partire dalle ultime tre generazioni e che nella popolazione attuale potremmo definire moderato. Tuttavia, i segni dell'isolamento sono ancora ben visibili nel genoma di quella parte della popolazione che è stata definita "nativa".

#### **4.3 - Quanto la genetica può dirci della storia di una popolazione?**

L'intuizione che la genetica sia uno strumento prezioso nella ricostruzione della storia non codificata sotto forma di scrittura, al pari dell'archeologia e della paleontologia, è un'intuizione non troppo recente, come mostrano i primi lavori di Cavalli-Sforza (Cavalli-Sforza et al. 1996)

Lo studio della popolazione di Campora è partito sulla base di alcune evidenze storiche. Tali evidenze sono state ampiamente confermate ed ampliate dallo studio del patrimonio genetico di questa popolazione. Il lavoro da me effettuato si può considerare come una indagine mirata a rispondere ad una domanda precisa, ma lascia aperte molte altre possibilità di indagare in modo più approfondito la storia demografica e le relazioni di questa popolazione con quelle circostanti.

## BIBLIOGRAFIA

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. *Nature* 437:1299-320

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-65

Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB, Fraumene C, Bebbere D, Forabosco P, Pirastu M (2001) Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. *Hum Genet* 109:198-209

Arcos-Burgos M, Muenke M (2002) Genetics of population isolates. *Clin Genet* 61:233-47

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, VILLEMS R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994-1004

Calcagno G, Labruna G, Sacchetti L (2005) Y-chromosome short tandem repeat (STR) haplotypes in a Campania population sample. *Clin Chem Lab Med* 43:163-6

Cann RL, Brown WM, Wilson AC (1982) Evolution of human mitochondrial DNA: a preliminary report. *Prog Clin Biol Res* 103 Pt A:157-65

Cann RL, Wilson AC (1983) Length mutations in human mitochondrial DNA. *Genetics* 104:699-711

Cavalli-Sforza LL, Menozzi P, Piazza A (1996) The history and geography of human genes. Princeton University Press, Princeton, N.J.

Cothran EG, MacCluer JW, Weitkamp LR, Pfennig DW, Boyce AJ (1984) Inbreeding and reproductive performance in standardbred horses. *J Hered* 75:220-4

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York,

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-9

Gianfrancesco F, Esposito T, Ombra MN, Forabosco P, Maninchedda G, Fattorini M, Casula S, Vaccargiu S, Casu G, Cardia F, Deiana I, Melis P, Falchi M, Pirastu M (2003) Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. *Am J Hum Genet* 72:1479-91

Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer Associates, Sunderland, Mass.

Helgason A, Nicholson G, Stefansson K, Donnelly P (2003) A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet* 67:281-97

Helgason A, Sigurethardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 67:697-717

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95-108

Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. *Genetics* 152:1711-22

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-13

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 62:1171-9

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22:139-44

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. *Nat Genet* 17:435-8

Laan M, Wiebe V, Khusnutdinova E, Remm M, Paabo S (2005) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. *Eur J Hum Genet* 13:452-62

Lewontin RC (1964) The Interaction of Selection and Linkage. Ii. Optimum Models. *Genetics* 50:757-82

Mayr E (1976) *Evolution and the diversity of life*. Belknap Press, Cambridge, Mass.

Palmer LJ, Cardon LR (2005) Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* 366:1223-34

Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8:1913-23

Peltonen L, McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* 291:1224-9

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. *Nat Rev Genet* 1:182-90

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* 411:199-204

Soodyall H, Jenkins T, Mukherjee A, du Toit E, Roberts DF, Stoneking M (1997) The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. *Am J Phys Anthropol* 104:157-66

Soodyall H, Nebel A, Morar B, Jenkins T (2003) Genealogy and genes: tracing the founding fathers of Tristan da Cunha. *Eur J Hum Genet* 11:705-9

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* 25:324-8

Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum Mol Genet* 13:25-33

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. *Nature* 394:138-40

Torrioni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835-50

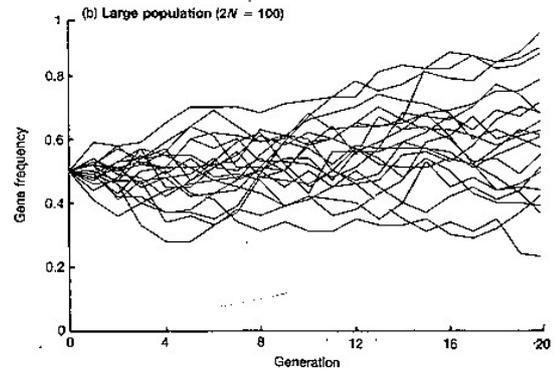
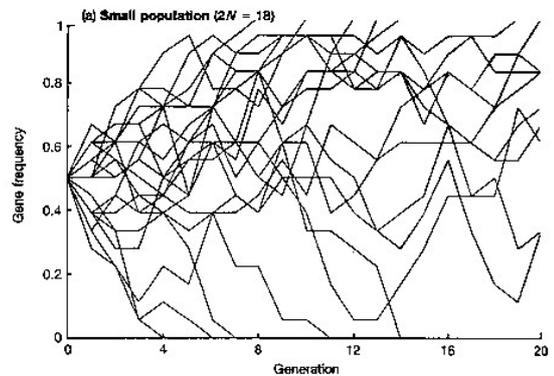
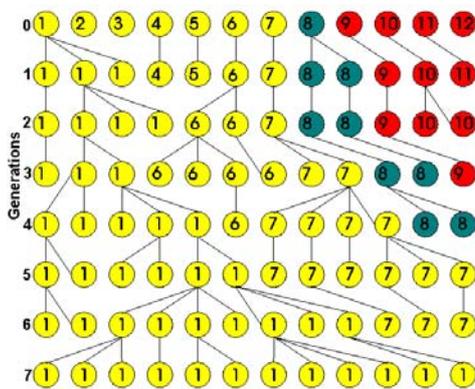
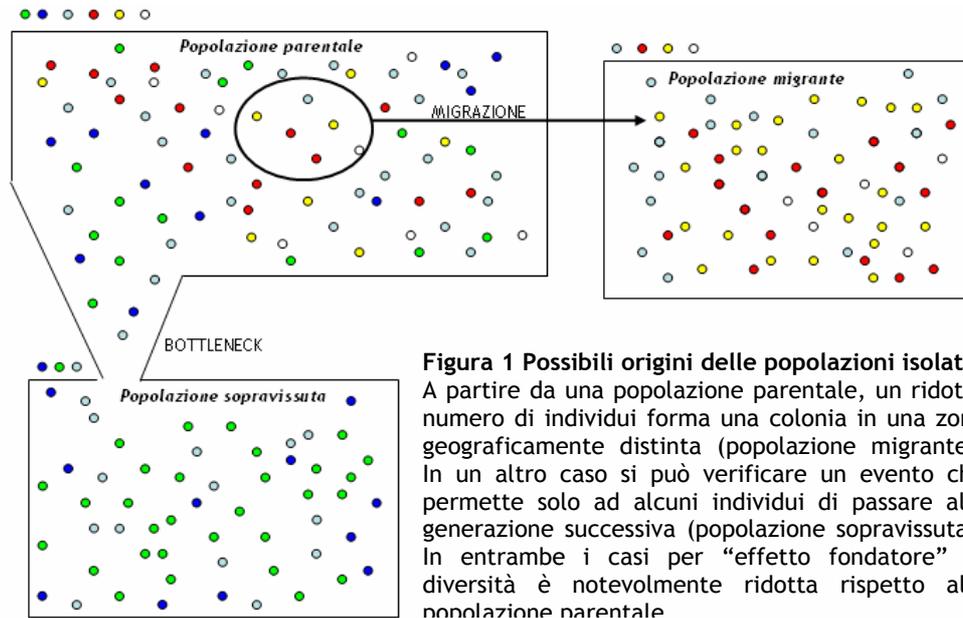
Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L (2000) Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. *Eur J Hum Genet* 8:604-12

Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. *Curr Opin Genet Dev* 14:316-23

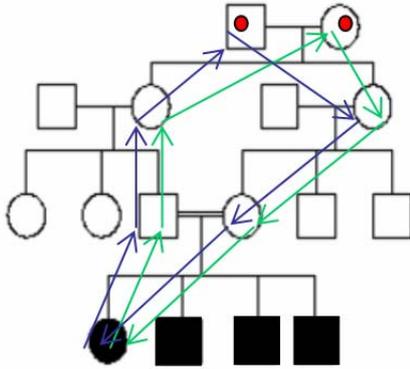
Weiss LA, Abney M, Cook EH, Jr., Ober C (2005) Sex-specific genetic architecture of whole blood serotonin levels. *Am J Hum Genet* 76:33-41

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. *Nat Genet* 23:397-404

Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Todd JA, Cucca F (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 9:2947-57



(Da: Cavalli-Sforza, Menozzi et al. 1996)



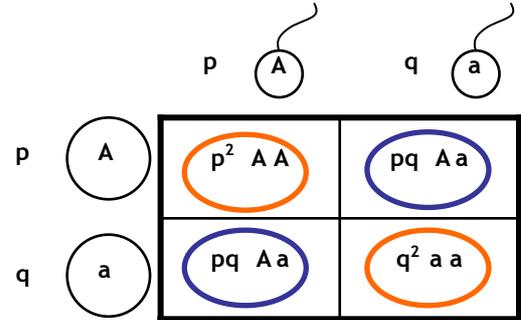
**Figura 3 Inbreeding.** Gli individui indicati in nero sono considerati “inbred” in quanto i loro genitori derivano da una stessa coppia di antenati (indicati da un pallino rosso). Le frecce in verde ed in blu, indicano i percorsi circolari (loops) che legano un individuo inbred ai due antenati comuni e ai suoi genitori.

Legge di Hardy-Weinberg:

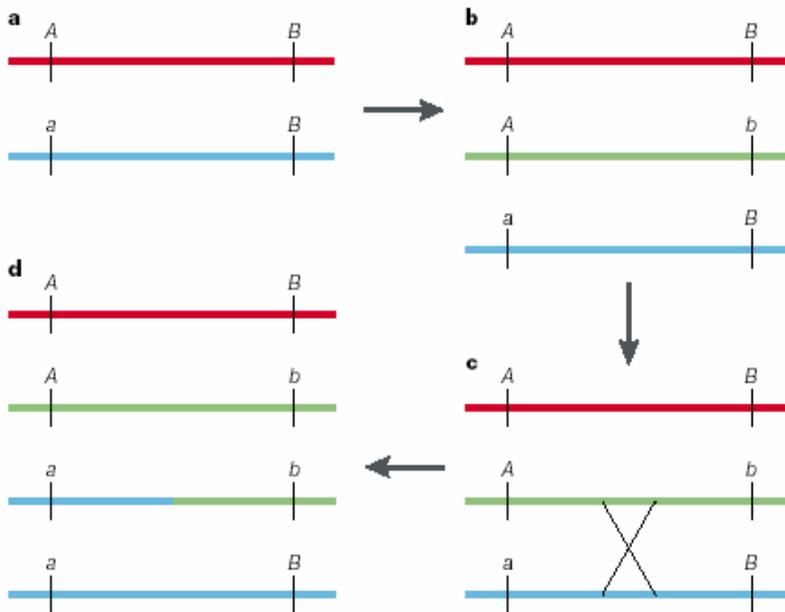
$$p^2 + 2pq + q^2 = 1$$

ETEROZIGOTI ATTESI =  $2pq$

OMOZIGOTI ATTESI =  $p^2 + q^2$



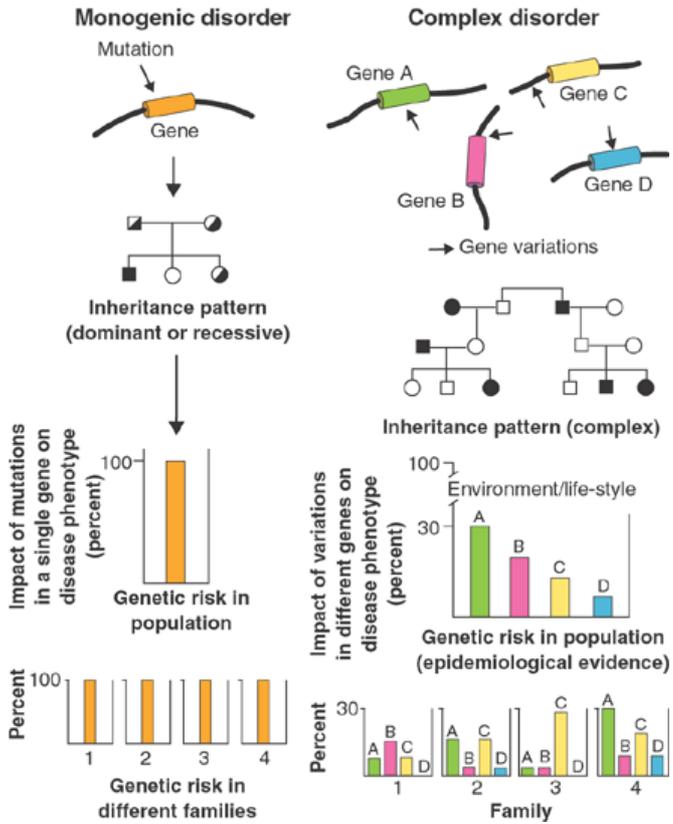
**Figura 4 Equilibrio di Hardy-Weinberg.** La legge di HW predice le frequenze dei genotipi attesi a partire dalle frequenze alleliche. Sebbene tale legge sia stata formulata per popolazioni ideali, è comunque possibile applicarla alle popolazioni reali



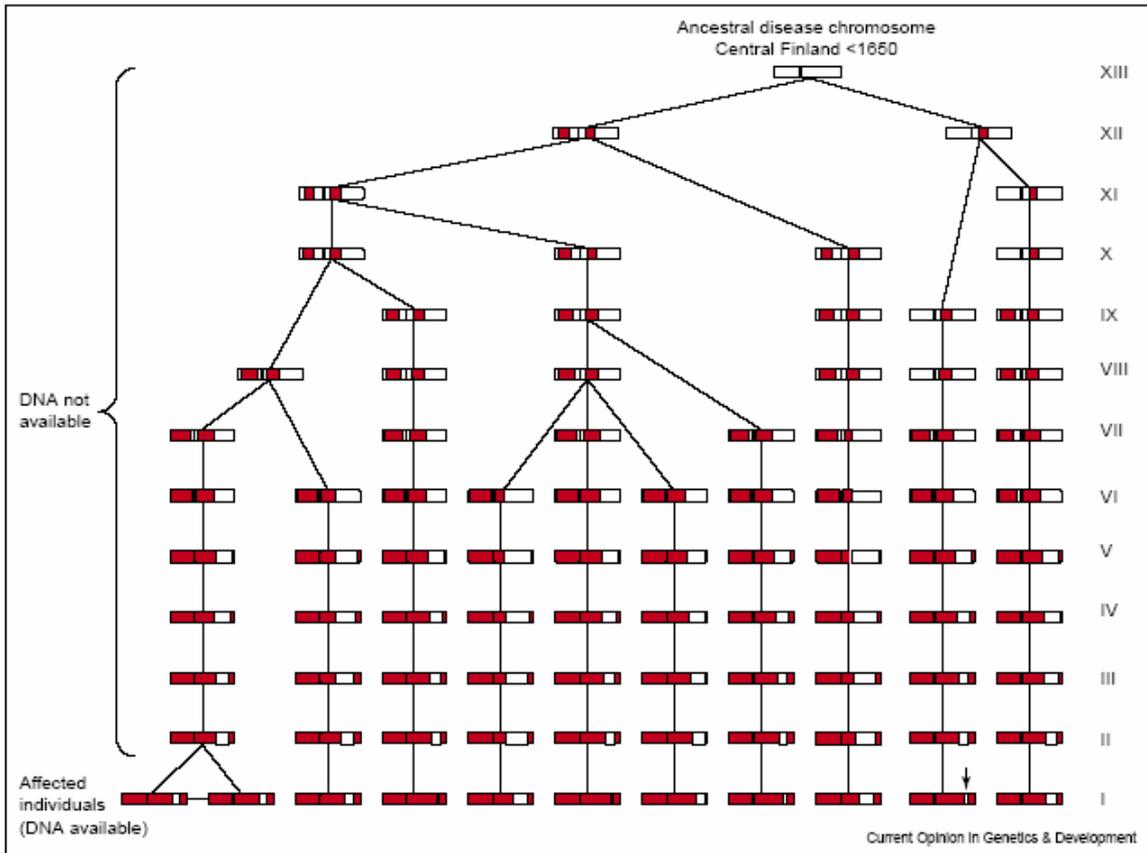
**Fig 5. Mutazione e ricombinazione modellano le regioni di Linkage Disequilibrium.** (a) Si consideri il locus polimorfico che presenta i due alleli: A ed a (b) Quando si verifica una mutazione al locus adiacente, che trasforma l'allele B in b, questa avviene su di un singolo cromosoma che porta un allele specifico (nell'esempio A). Quindi, subito dopo l'insorgenza della mutazione, sarà possibile osservare nella popolazione, solo tre dei quattro possibili aplotipi. Inoltre l'allele b sarà sempre associato all'allele A. Col tempo, l'associazione tra questi due alleli verrà gradualmente erosa dalla ricombinazione tra i loc. (d) Questo si traduce nella generazione del quarto aplotipo e nel declino del LD tra i markers associato all'incremento della frequenza dell'aplotipo ricombinante.

(Da: Ardlie, Kruglyak et al. 2002)

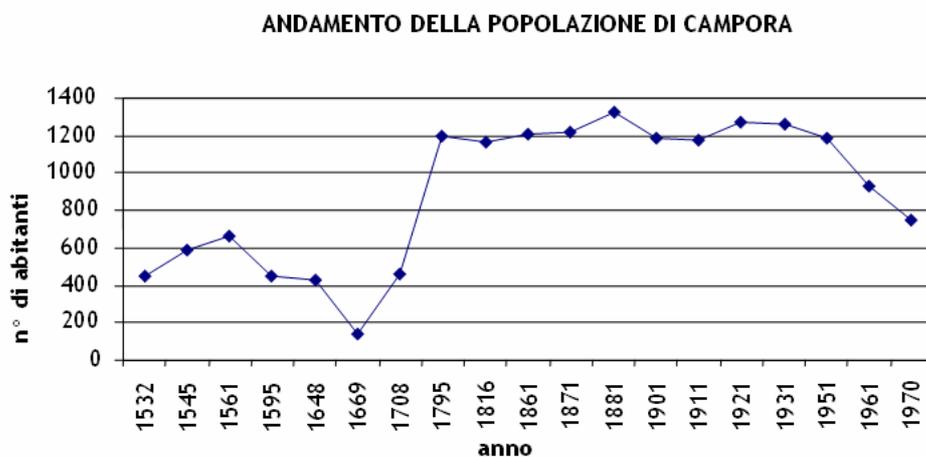
**Fig6. Tratti mendeliani classici e tratti complessi**



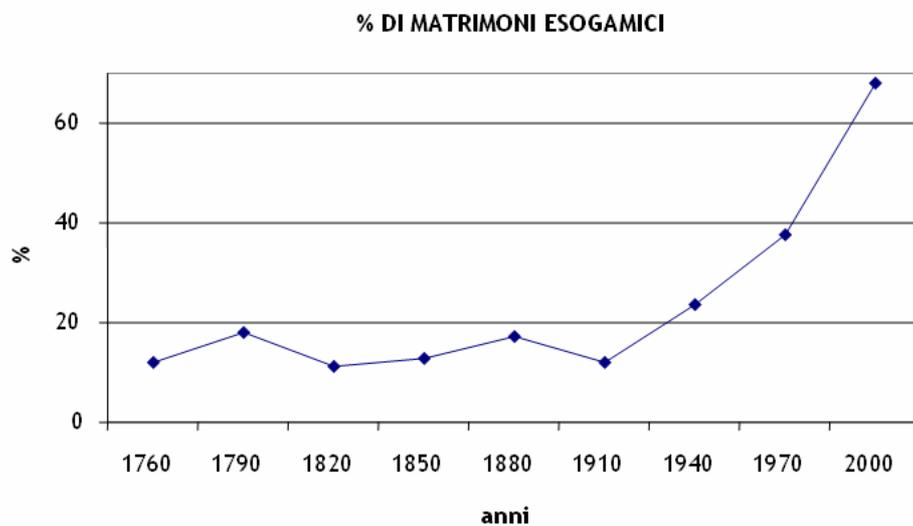
(D : Peltone and McKusick 2001 )



**Fig 7. Uso del Linkage Disequilibrium nella ricerca dei geni nelle popolazioni isolate con un ridotto numero di fondatori.** La ricombinazione in 13 generazioni ha delimitato la regione che contiene i geni responsabili (in bianco). Comparando il genoma degli individui affetti si può identificare questa regione condivisa ed utilizzarla per ulteriori analisi



**Figura8.** Numero di abitanti per anni in base a quanto riportato dallo storico P. Ebner. Fino al 1700 i dati derivano da fonti ecclesiastiche e riportano anche il numero di fuochi ovvero di nuclei familiari pluri-generazionali



**Figura 9.** Matrimoni esogamici per generazione secondo i registri di matrimonio.

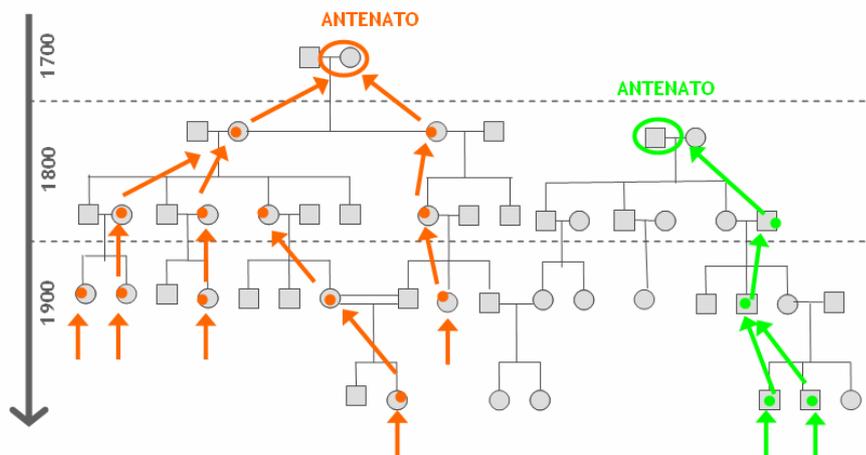
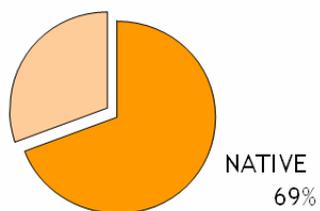


Figura 10 Ricostruzione delle linee genealogiche.

**POPOLAZIONE FEMMINILE VIVENTE**

NON NATIVE  
31%



**POPOLAZIONE MASCHILE VIVENTE**

NON NATIVI  
24%

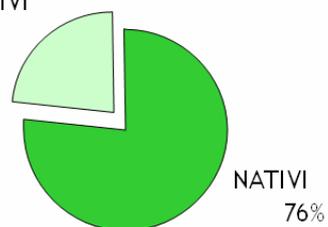


Figura 11. Calcolo della popolazione nativa di Campora rispetto alla popolazione anagrafica. Nei grafici è rappresentata la percentuale di individui che sono nativi di Campora, in quanto discendenti da capostipiti che sono arrivati nel paese al massimo tre generazioni fa.

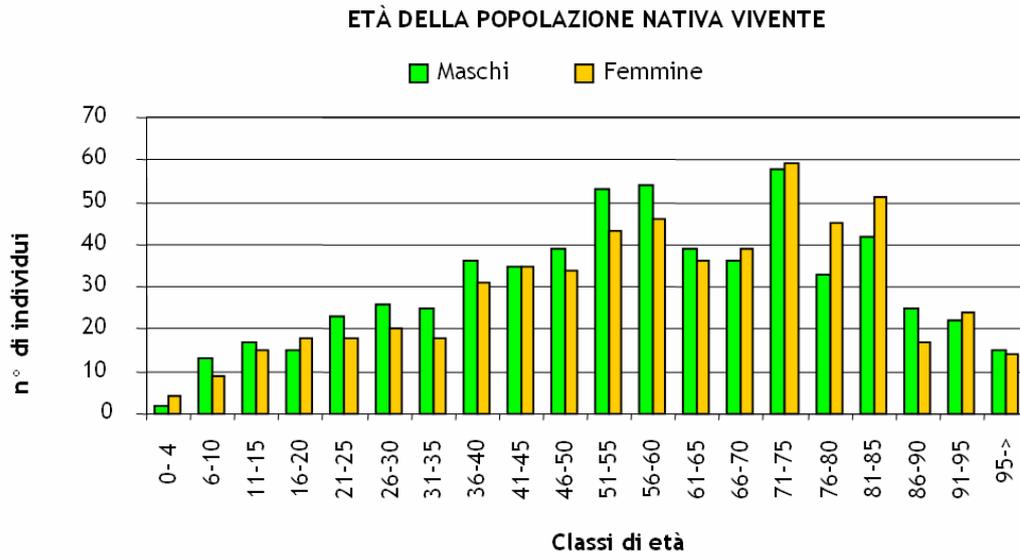


Fig12. Classi di età della popolazione vivente

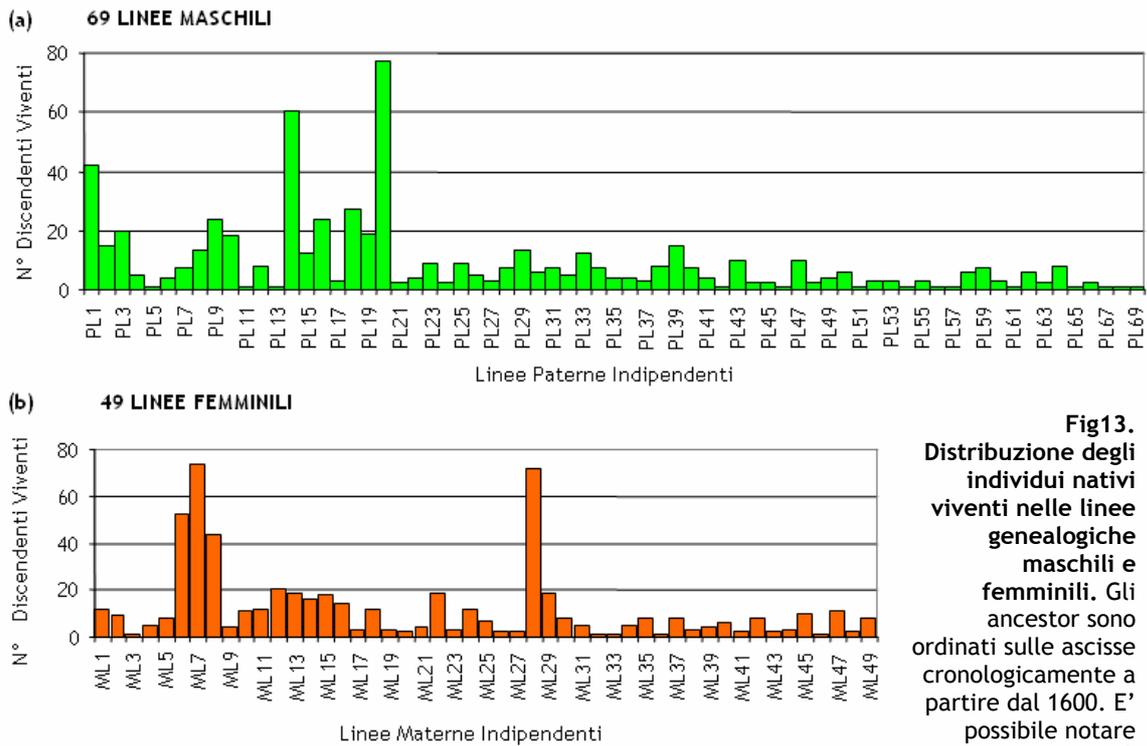


Fig13. Distribuzione degli individui nativi viventi nelle linee genealogiche maschili e femminili. Gli ancestor sono ordinati sulle ascisse cronologicamente a partire dal 1600. E' possibile notare l'estrema variabilità nella discendenza di ciascun ancestor

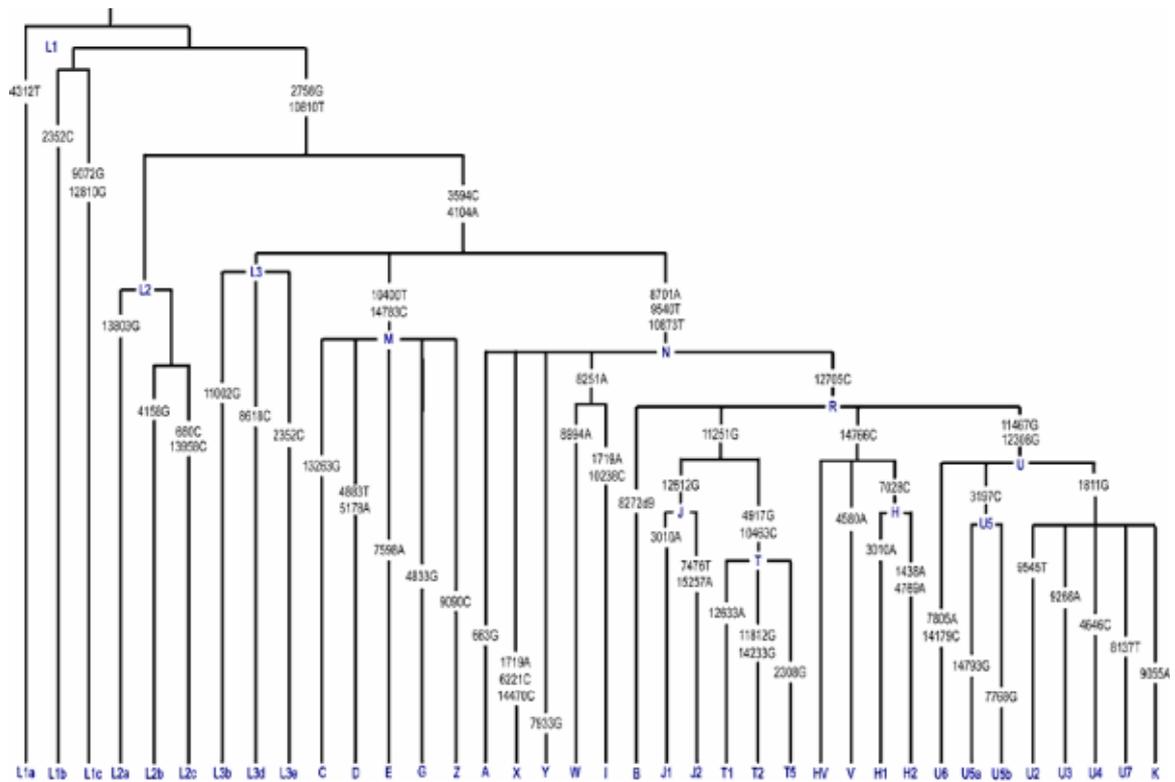


Fig14. Classificazione filogenetica degli aplogruppi mitocondriali. (da mtDB - Human Mitochondrial Database, <http://www.genpat.uu.se/mtDB/>)

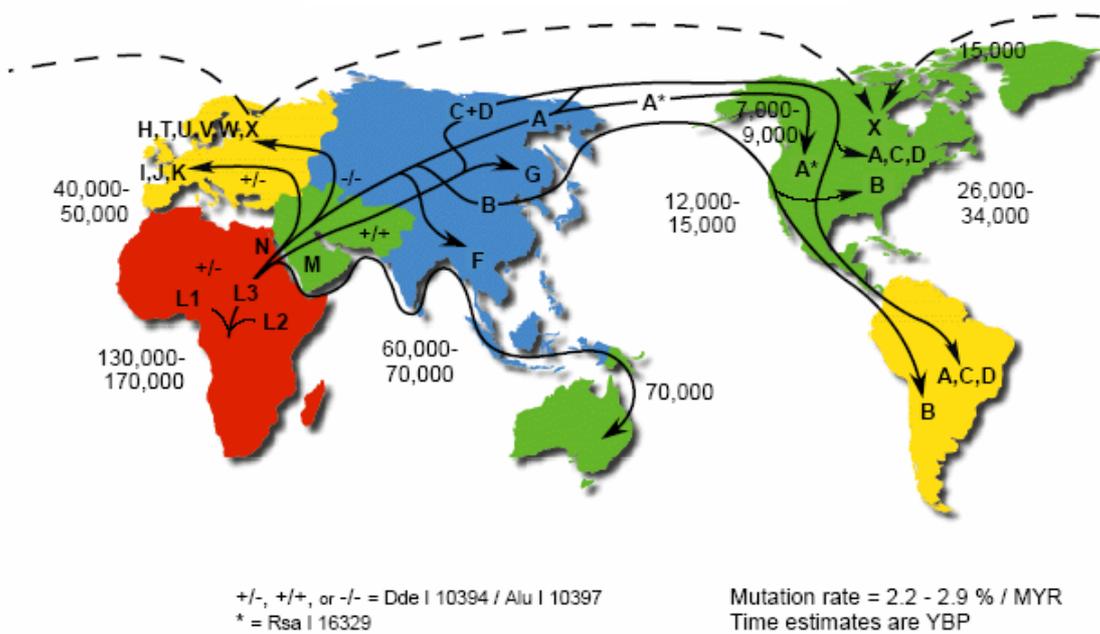
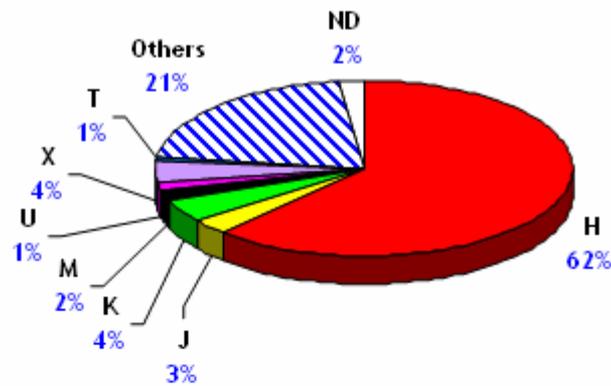


Fig15. Origine, migrazione e distribuzione dei principali aplogruppi mitocondriali. (da MITOMAP A human mitochondrial genome database, <http://www.mitomap.org/>)

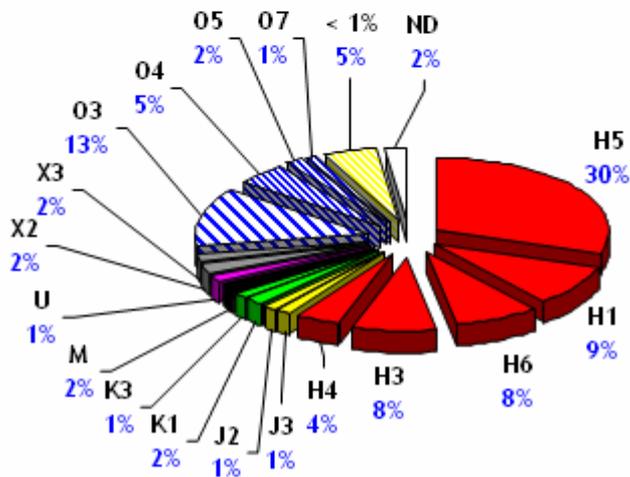
a



**Fig16. Linee fondatrici femminili.**

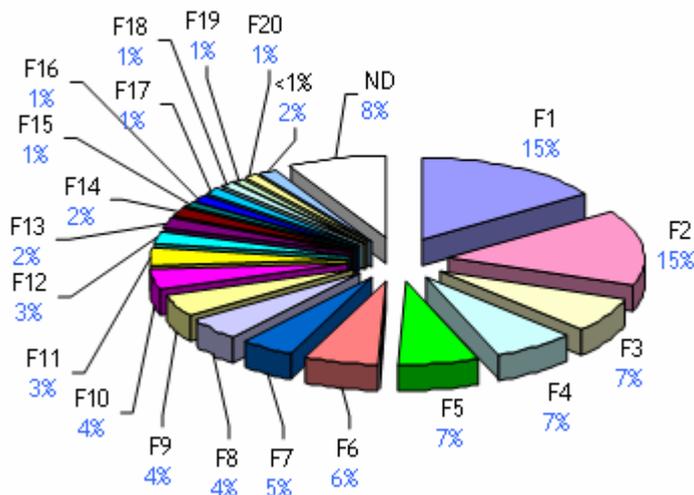
(a) Risultato dell'analisi di restrizione. Ogni lettera corrisponde ad un aplogruppo, mentre la percentuale si riferisce alle femmine native viventi che rappresenta. Questo tipo di analisi si è rivelata non sufficientemente discriminante, in quanto la maggior parte delle femmine, come atteso appartiene all'aplogruppo H, maggiormente diffuso in Europa, ed non è stato possibile classificare il 21% della popolazione.

b

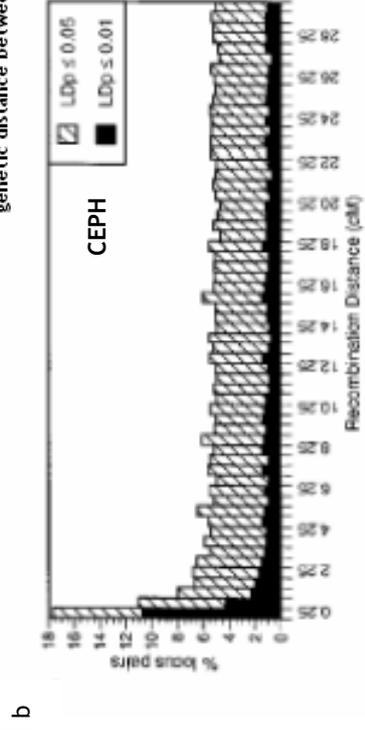
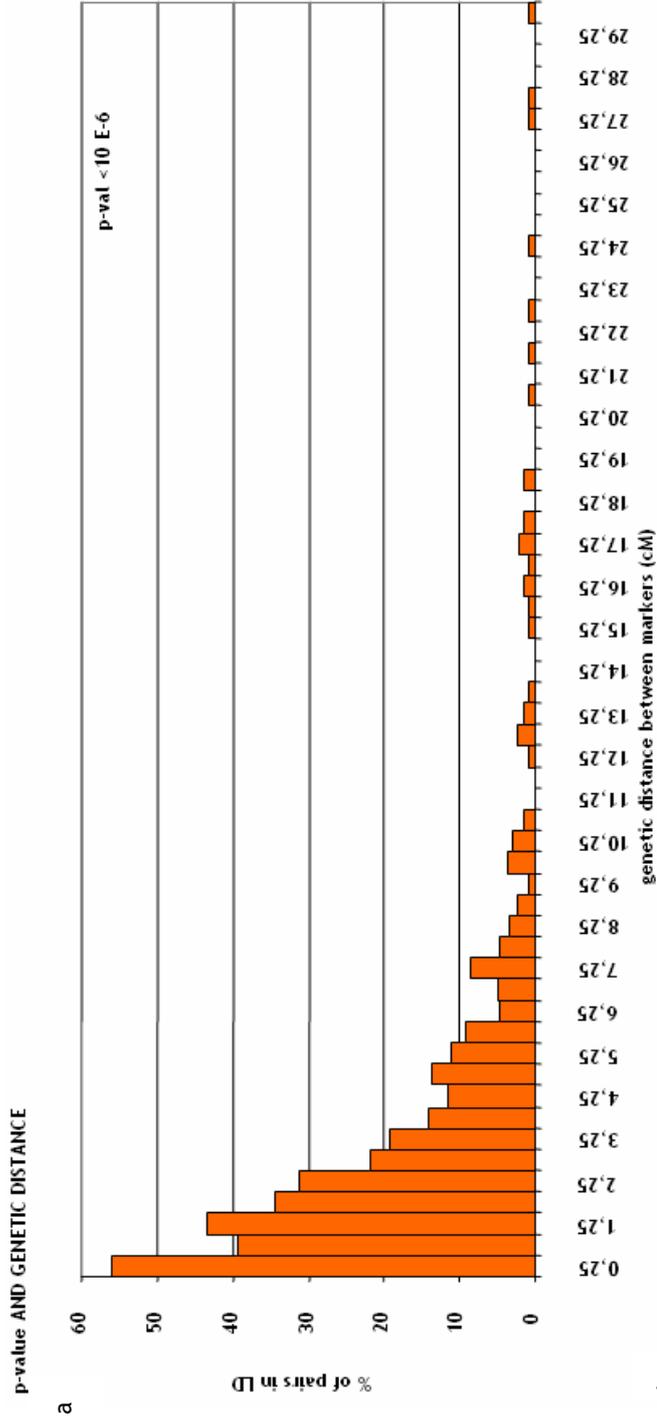


In fig 16b, ogni fetta della torta corrisponde ad una linea fondatrice. Per ogni linea fondatrice la lettera grande indica l'aplogruppo di restrizione; il numero rappresenta il sotto-aplogruppo in base alla classificazione effettuata con i polimorfismi della regione ipervariabile; la percentuale si riferisce alle femmine native viventi che rappresenta.

Dodici fondatrici hanno contribuito con meno dell' 1% alla popolazione attuale, e sono stati raggruppati in una singola categoria denominata "<1%". Inoltre, un 2% delle femmine native viventi non sono state classificate perchè non è stato possibile effettuare il campionamento.



**Fig17. Linee fondatrici maschili.** Per ciascuna linea è indicata la percentuale di sub-popolazione maschile che rappresenta.



**Figure 18**

(Da: G. A. Huttley et al. Genetics 1999)

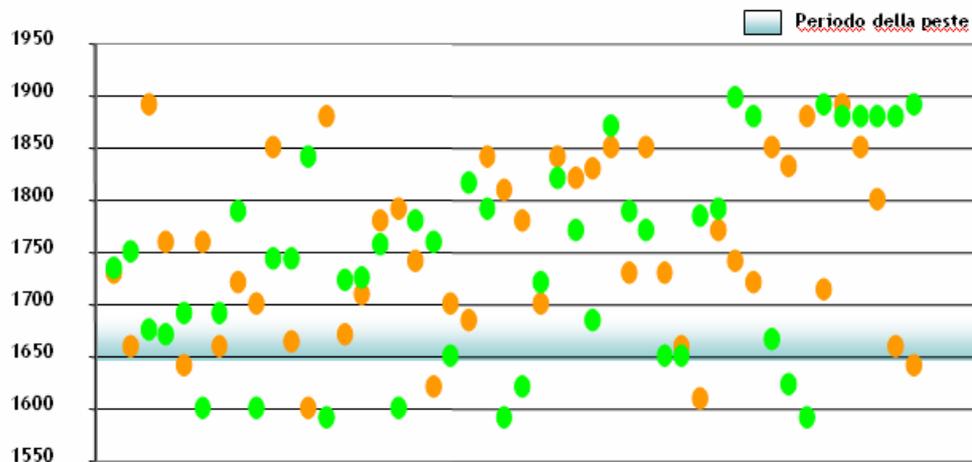
>													
1	7823566	7823703	7823751	m	n	3	5	4	5	2	5	7	7
1	7823528	7823703	7823751	f	n	7	7	2	6	4	5	5	10
1	7823703	0	0	m	y	3	7	2	5	4	5	5	7
1	7823751	0	0	f	y	5	7	4	6	2	5	7	10
2	7823428	7823717	7825890	m	n	4	7	2	2	2	4	4	9
2	7823443	7823717	7825890	m	n	1	7	x	x	2	2	4	4
2	7825890	0	0	f	y	1	4	2	6	2	4	4	9
2	7823717	0	0	m	y	7	7	x	x	2	2	4	4
3	7823419	7823723	7823827	m	n	1	7	2	2	2	6	4	7
3	7823521	7823723	7823827	m	n	7	8	2	4	2	2	4	7
3	7823723	0	0	m	y	7	7	2	4	2	6	4	4
3	7823827	0	0	f	y	1	8	x	x	2	5	7	7
4	7823555	7823734	7823800	m	n	4	4	2	2	2	4	5	11
4	7823538	7823734	7823800	f	n	7	8	3	7	2	2	4	4
4	7823734	0	0	m	y	4	7	2	3	x	x	4	11
4	7823800	0	0	f	y	4	8	2	7	2	4	4	5

Figura 19 Esempio di un file in formato linkage. Nella prima colonna è indicato il pedigree di appartenenza, a seguire l'ID individuo, l'ID del padre, l'ID della madre, il sesso e lo stato. A seguire i marcatori

PCR fragment										Haplogroup
1	2	3	4 II	4 III	5	6 A	6 D	7	8	
Sites (a)										
- Hinf I 18096	- Sde I 1715	- Aiu I 7025	- Uae II 9062	- Uae III 8994	+ G 18308	+ Aiu I 10397	+ Sde I 10394	+ Aiu I 10028	+ Aiu I 15806	
-	-	+	-	-	-	-	-	-	-	H
-	+	-	-	-	-	-	+	+	-	I
+	-	-	-	-	-	-	+	-	-	J
-	-	-	+	-	+	-	+	-	-	K
-	-	-	-	-	-	-	-	-	+	T
-	-	-	-	-	+	-	-	-	-	U
-	-	-	-	-	-	-	-	-	-	V
-	-	-	-	+	-	-	-	-	-	W
-	+	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	+	+	-	-	M
-	-	-	-	-	-	-	-	-	-	Other

Figura 20

(a) restriction sites are listed as either absent (-) or present (+) relative to the reference sequence GENE BANK accession NC\_001807 (Anderson et al. 1981 review by Ingman M., Kaessmann H., Paabo S. and Crallansten U. Nature 408 6813: 708-713 (2000)) map position of sites were also confirmed using [hftcmap](http://www.hftcmap.org) databases ([www.hftcmap.org](http://www.hftcmap.org))  
Haplogroup reference:- Torroni et al. Genetic 1996.



	% di individui nativi viventi derivanti da un antenato vissuto prima o durante la peste
femmine	82
maschi	74

Figura 21. Posizione cronologica degli ancestor rispetto al periodo della peste. Nel grafico gli ancestor maschi (verdi) e femmine (arancio) sono posizionati lungo le ordinate in base alla loro data di nascita, mentre sono (in ordine del tutto casuale sulle ascisse)

Tabella 1. Alcuni esempi di popolazioni isolate.

Motivi dell'isolamento	Popolazione
Geografico	Palau, Tristan da Chuna, Sardegna
Culturale	Amish, Hutterites
Geografico & Culturale	Jewish, Basques, Finnish

Tabella 2. Distribuzione dei capostipiti nel

	capostipiti femmine		capostipiti maschi	
	totali	con discendenti viventi	totali	con discendenti viventi
1600 -1700	39	13	22	20
1700 -1800	111	18	23	21
1800 -1900	59	18	37	28
<b>totale 1600-1900</b>	<b>209</b>	<b>49</b>	<b>82</b>	<b>69</b>
1900 -2000	101	97	165	147
<b>totale 1600-2000</b>	<b>310</b>	<b>146</b>	<b>247</b>	<b>216</b>

Tab3. Valori del coefficiente di inbreeding

	<b>nativi viventi</b>	<b>% con f≠0</b>	<b>f</b>	<b>SD</b>
maschi	608	77	0,0055	0,0074
femmine	576	87	0,0075	0,0106
<b>valore medio nella popolazione</b>			<b>0,0065</b>	<b>0,0091</b>

Nome	Indice	16017	16037	16069	16086	16093	16104	16111	16126	16129	16136	16145	16146	16148	16153	16163	16172	16183	16184	16186	16187	16189	16192	16193	16194	16224	16225	16244	16249	16250	16257	16261	16262	16279	16285	16287	16294	16295	16299				
h5	29,9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
o3	12,8	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
h1	9,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h6	8,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h3	7,6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
o4	4,9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h4	4,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
m	2,1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
x2	2,1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
x3	2,1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
k1	1,9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
o5	1,7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
j3	1,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
x3	1,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
u	1,4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
j2	1,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
o7	1,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
k2	0,9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h7	0,7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
l1	0,7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
h2	0,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
h8	0,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
h9	0,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
h10	0,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
j1	0,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
o1	0,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
h2	0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NC 001807		T	A	C	T	T	C	C	T	G	T	G	A	C	G	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T		

**Tabella 4 Polimorfismi del DNA mitocondriale.** I numeri nella prima riga indicano la posizione sul DNA mitocondriale secondo la sequenza di riferimento (aggiungere rif), mentre le lettere nell'ultima riga indicano la sequenza di riferimento nella posizione specificata. Nella prima colonna in ciascuna riga vi è una linea fondatrice femminile identificata da una lettera che specifica l'aplogruppo di restrizione ed un numero che specifica il sotto-aplogruppo. la seconda colonna indica il percento di

**Tabella 6. Aplogruppi presenti sul cromosoma Y.** Nella prima colonna in ciascuna riga è riportata un linea fondatrice diversa e nella seconda colonna la percentuale di sub-popolazione maschile che rappresenta. Nelle colonne successive sono mostrati i polimorfismi che caratterizzano ciascun aplogruppo.

<b>founder</b>	<b>% pop</b>	<b>DYS393</b>	<b>DYS19</b>	<b>DYS389</b>	<b>DYS390</b>	<b>DYS391</b>	<b>DYS392</b>	<b>DYS385</b>
<b>1</b>	15,46	120	192	255	220	290	258	360
<b>2</b>	14,97	124	188	255	220	286	252	380
<b>3</b>	6,91	124	196	259	224	286	252	360
<b>4</b>	6,58	124	188	259	220	282	252	368
<b>5</b>	6,58	124	192	255	220	290	258	360
<b>6</b>	6,09	128	196	255	212	290	252	368
<b>7</b>	4,61	120	192	255	220	286	252	364
<b>8</b>	4,44	120	192	255	224	290	252	368
<b>9</b>	4,44	128	188	255	220	286	252	380
<b>10</b>	3,95	120	196	251	220	286	252	372
<b>11</b>	3,13	120	192	251	212	286	252	372
<b>12</b>	2,63	120	196	255	220	290	258	360
<b>13</b>	1,81	124	200	251	208	286	252	372
<b>14</b>	1,64	120	196	255	216	282	252	372
<b>15</b>	1,32	124	200	259	224	290	252	360
<b>16</b>	1,32	124	196	255	220	290	258	360
<b>17</b>	1,32	124	196	255	220	290	252	372
<b>18</b>	1,15	124	192	259	220	290	258	360
<b>19</b>	0,99	120	192	259	224	286	258	368
<b>20</b>	0,99	124	192	255	216	282	252	380
<b>21</b>	0,66	120	192	255	224	286	252	372
<b>22</b>	0,49	124	200	255	220	286	252	372
<b>23</b>	0,33	124	188	255	220	290	258	364
<b>24</b>	0,16	124	200	255	220	290	258	360

Tabella 5. Numero di alleli dei microsatelliti utilizzati per la caratterizzazione degli aplogruppi nella sub-popolazione maschile.

STR	n° di Alleli
DYS19	12
DYS389	11
DYS390	14
DYS391	10
DYS392	13
DYS393	11
DYS385	>100

Tabella 7. Caratteristiche dei markers del cromosoma X utilizzati nel calcolo del Linkage Disequilibrium

marker	Posizione sulla mappa (Mb) §	Eterozigosità (Campora)	Eterozigosità (GDB°)
DXS983	68,31	0,73	0,63
DXS8092	72,99	0,85	0,87
DXS8037	72,99	0,69	0,74
DXS1225	76,97	0,76	0,74
DXS8082	77,13	0,82	0,74
DXS986	78,14	0,68	0,76

° <http://gdbwww.gdb.org/>

§ (UCSC Genome Browser, March 2004)

MARKER PAIRS	MARKERS DISTANCE		SAAMI $\Upsilon$ N= 54	GAVDI $\circ$ N= 73	Campora N= 53	SWEDEN $\Upsilon$ N= 41	SARDINIA $\circ$ N= 73	UK $\circ$ N= 73	FINLAND $\Upsilon$ N= 80	ESTONIA $\Upsilon$ N= 45
	Mb $\S$	cM $\Upsilon$								
DXS8092	0,00	0,40	0,000	0,000	0,045	0,028	0,280	0,620	0,180	0,072
DXS8092	1,01	0,20	0,000	0,000	0,001	0,618	0,322	0,884	0,092	0,143
DXS1225	1,17	0,50	0,000	0,000	0,104	0,448	0,166	0,703	0,393	0,688
DXS1225	1,62	0,30	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
DXS8037	3,98	0,00	0,091	0,008	0,009	0,242	0,710	0,647	0,836	0,488
DXS8092	3,98	0,40	0,000	0,000	0,008	0,676	0,921	0,320	0,283	0,120
DXS8037	4,14	0,30	0,012	0,004	0,002	0,033	0,630	0,002	0,238	0,625
DXS8092	4,14	0,10	0,000	0,000	0,001	0,102	0,319	0,492	0,044	0,065
DXS983	4,64	1,60	0,000	0,000	0,000	0,746	0,876	0,974	0,314	0,153
DXS983	4,68	2,00	0,300	0,000	0,407	0,924	0,036	0,149	0,683	0,104
DXS8037	5,15	0,50	0,000	0,003	0,119	0,256	0,302	0,975	0,620	0,739
DXS8092	5,15	0,10	0,000	0,000	0,006	0,332	0,125	0,940	0,331	0,100
DXS983	8,65	2,00	0,000	0,170	0,001	0,480	0,169	0,338	0,630	0,520
DXS983	8,82	1,70	0,000	0,245	0,050	0,082	0,142	0,243	0,565	0,730
DXS983	9,82	1,50	0,000	0,003	0,042	0,400	0,825	0,253	0,829	0,468

significativo (p<0.05)

non significativo (p≥0.05)

$\circ$  Zavattari et al. 2001

$\Upsilon$  (Lauri & Paabo, 1997)

$\S$  (UCSC Genome Browser, March 2004)

Tabella8. LD nelle regioni Xq13.3Xq21.1

Tabella 9. Descrizione dei marcatori genotipizzati nella popolazione di Campora. Per ogni cromosoma è indicato il numero di marcatori, la distanza genetica media tra marcatori sul cromosoma secondo la mappa genetica della DeCode (<http://www.decode.com/>), la distanza fisica media, il numero di possibili accoppiamenti tra marcatori, la soglia reale considerata nei calcoli di Linkage Disequilibrium

<b>cromosoma</b>	<b>n° di markers</b>	<b>spaziatura media (cM)</b>	<b>spaziatura media (Mb)</b>	<b>possibili coppie</b>	<b><math>\alpha</math> reale (per <math>\alpha=0,05</math>)</b>
1	77	3,47	2,72	2926	1,71E-05
2	79	3,38	3,10	3081	1,62E-05
3	65	3,5	3,10	2080	2,40E-05
4	58	3,66	3,33	1653	3,02E-05
5	62	3,44	2,94	1891	2,64E-05
6	56	3,43	3,06	1540	3,25E-05
7	59	3,25	2,71	1711	2,92E-05
8	52	3,07	2,71	1326	3,77E-05
9	51	3,29	2,79	1275	3,92E-05
10	53	3,51	2,59	1378	3,63E-05
11	46	3,43	3,06	1035	4,83E-05
12	51	3,5	2,65	1275	3,92E-05
13	32	4,24	3,02	496	1,01E-04
14	44	2,83	1,95	946	5,29E-05
15	40	3,19	1,91	780	6,41E-05
16	36	3,52	2,55	630	7,94E-05
17	41	3,32	2,00	820	6,10E-05
18	35	3,35	2,15	595	8,40E-05
19	33	3,56	1,92	528	9,47E-05
20	32	3,16	1,43	496	1,01E-04
21	20	2,64	1,38	190	2,63E-04
22	22	3,11	1,44	231	2,16E-04

Tabella 10

<i>PCR fragment</i>										<i>Haplogroup</i>
1	2	3	4 II	4 III	5	6 A	6 D	7	8	
<i>Sites (a)</i>										
- <i>Hinf</i> I 16065	- <i>Dde</i> I 1715	- <i>Alu</i> I 7025	- <i>Hae</i> II 9062	- <i>Hae</i> III 8994	+ <i>G</i> 16308	+ <i>Alu</i> I 10397	+ <i>Dde</i> I 10394	+ <i>Alu</i> I 10028	+ <i>Alu</i> I 15605	
-	-	+	-	-	-	-	-	-	-	H
-	+	-	-	-	-	-	+	+	-	I
+	-	-	-	-	-	-	+	-	-	J
-	-	-	+	-	+	-	+	-	-	K
-	-	-	-	-	-	-	-	-	+	T
-	-	-	-	-	+	-	-	-	-	U
-	-	-	-	-	-	-	-	-	-	V
-	-	-	-	+	-	-	-	-	-	W
-	+	-	-	-	-	-	-	-	-	X
-	-	-	-	-	-	+	+	-	-	M
-	-	-	-	-	-	-	-	-	-	<u>Other</u>

(a) restriction sites are listed as either absent (-) or present (+) relative to the reference sequence GENE BANK accession NC\_001807 (Anderson *et al.* 1981 review by Ingman M., Kaessmann H., Paabo S. and Gyllenstein J. *Nature* 408 6813: 708-713 (2000)) map position of sites were also confirmed using [Mitomap](http://www.mitomap.org) databases

([www.mitomap.org](http://www.mitomap.org))

Haplogroup reference: [Torroni et al. Genetic](#) 1996.

Tabella 11

	n° di antenati	n° medio di discendenti viventi	standard deviation	coefficient of variation
femmine	46	5,04	19,62	3,89
maschi	69	8,69	12,81	1,47

Tab 11 *Haplotype diversity* sul cromosoma Y 1n alcune popolazioni europee

POPULATION	GEOGRAPHIC PROVENIENCE	N	k	HD	RANK
Finnish	Most regions	280	54	<b>0,873</b>	16
Scottish	South-West	61	26	<b>0,916</b>	15
Icelandic	Most regions	181	50	<b>0,921</b>	14
Ashkenazic Jewish	Northern Europe	170	55	<b>0,938</b>	13
<b>Campora</b>	<b>South-Italy</b>	<b>87</b>	<b>24</b>	<b>0,944</b>	<b>10</b>
Norwegian	Most regions	412	98	<b>0,957</b>	8
Swedish	North & Gotland	112	46	<b>0,958</b>	7
Italian	North & Rome	555	192	<b>0,975</b>	3
Swiss	Bern & Lousanne	199	96	<b>0,975</b>	2
Hungarian	Budapest	117	60	<b>0,982</b>	1

