



UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"Biologia Evoluzionistica ed Ecologia"

CICLO XXXVI

*Pompei, capsula del tempo dell'Impero Romano:
analisi paleogenomica dei resti umani rinvenuti nell'antica città*

Coordinatore:
Prof. Pierluigi Viaroli

Tutore:
Prof. David Caramelli

Co-tutore:
Prof. Guido Barbujani

Dottorando: Dott.ssa Costanza Cannariato

Anni Accademici 2020/2021 – 2022/2023

Il dottorato in Biologia Evoluzionistica ed Ecologia – XXXVI ciclo è in convenzione tra le Università di Ferrara, Firenze e Parma (sede amministrativa).

Ad Annuccia.

SOMMARIO

ABSTRACT	6
SUMMARY	7
1 INTRODUZIONE.....	9
1.1 Pompei.....	9
1.1.1 Primi studi sul DNA antico dai resti umani di Pompei	12
1.2 Genetica delle popolazioni Europee.....	14
1.2.1 L'Italia Imperiale dal punto di vista genomico	17
1.3 Il DNA antico	19
1.3.1 Tecniche di analisi dell'aDNA	22
2 OBIETTIVO	24
3 MATERIALI E METODI	26
3.1 Materiale Archeologico	26
3.2 Caso studio: La Stanza degli Scheletri.....	29
3.3 Metodi di laboratorio	31
3.4 Analisi bioinformatica.....	34
3.4.1 Trattamento dei dati grezzi – <i>shotgun</i> esplorativo	34
3.4.2 Trattamento dei dati grezzi – <i>shotgun</i> profondo	35
3.4.3 Determinazione del sesso	36
3.4.4 Stima delle contaminazioni	36
3.4.5 Identificazione degli aplogruppi uniparentali.....	36
3.4.6 <i>Genotyping</i> e Analisi delle Componenti Principali	37
3.4.7 Analisi delle parentele	38
3.4.8 Approccio <i>Whole Genome</i> e <i>Genotype Likelihood</i>	38
3.4.8.1 Selezione degli individui di confronto	39
3.4.8.2 PCAngsd.....	40
3.4.8.3 NgsAdmix.....	40
4 RISULTATI E DISCUSSIONI.....	42
4.1 <i>Shotgun</i> esplorativo.....	42
4.1.1 Determinazione del sesso	51
4.1.2 <i>Genotyping</i> e analisi delle componenti principali (PCA)	54
4.2 <i>Shotgun</i> profondo.....	58
4.2.1 Determinazione del sesso	60
4.2.2 Stima delle contaminazioni	61
4.2.3 Identificazione degli aplogruppi uniparentali.....	62

4.2.4 <i>Genotyping</i> e Analisi delle Componenti Principali (PCA).....	63
4.2.5 Approccio <i>Whole Genome</i>	65
4.2.5.1 PCAngsd.....	66
4.2.5.2 NGSAdmix.....	67
4.3 La Stanza degli Scheletri	70
5 CONCLUSIONI	75
Bibliografia.....	77

ABSTRACT

La peculiare storia che ha vissuto l'antica città di Pompei ha fatto sì che essa abbia oggi un'importanza senza eguali in tutto il mondo. Per questo motivo il Parco Archeologico di Pompei è stato incluso all'interno dei patrimoni UNESCO. Gli scavi, già a partire dal XVIII secolo, hanno portato alla luce non solo ben preservati edifici pubblici e privati, ma anche numerosi resti umani riferibili alle vittime dell'eruzione del complesso Somma-Vesuvio avvenuta nel 79 e.c. Lo studio antropologico dei resti delle vittime è stato accompagnato, fin dalla metà degli anni '90, dall'analisi del loro DNA limitato però alle tecniche in uso in quel periodo. Grazie alla convenzione firmata con il Laboratorio di Ricerche Applicate di Pompei, nel laboratorio di Antropologia Molecolare e Paleogenomica dell'Università degli Studi di Firenze sono stati campionati per questo progetto di ricerca più di 170 reperti, in maggioranza ossa e denti, allo scopo di analizzare il genoma degli individui che vivevano o si trovavano in quel preciso periodo a Pompei. Grazie all'applicazione di protocolli specifici per il recupero di molecole di DNA degradate e all'impiego delle più sofisticate tecnologie di sequenziamento NGS (*Next Generation Sequencing*), è stata rilevata un'ottima preservazione del DNA endogeno che, in alcuni casi, supera anche il 90%. Con un approccio multidisciplinare è stato studiato un ristretto gruppo di individui, alcuni dei quali legati da una relazione di parentela, che hanno perso la vita all'interno della stanza di una *domus* Pompeiana a causa del crollo del tetto e dell'inalazione dei fumi vulcanici. Per la prima volta è stata studiata la variabilità genomica di un esteso campione della popolazione a Pompei ricostruendo il genoma di 12 individui con una profondità media di sequenziamento 1x. La ricostruzione di questi genomi a basso *coverage* ha spinto verso l'applicazione di metodi di analisi che considerano l'incertezza associata ai genotipi piuttosto che la loro chiamata diretta. Dall'analisi delle componenti principali (PCA) e dell'*Admixture* emerge in città la presenza di una grande variabilità genomica se si confronta con quella delle altre popolazioni già studiate sia di epoche precedenti nello stesso territorio, sia coeve in più punti del continente Europeo. Infatti, a Pompei si ritrova un insieme di varianti genetiche identificate in precedenza non solo nella penisola Italiana, ma anche al di fuori di essa. I risultati delle analisi genomiche trovano ampio supporto dalle testimonianze storiche che vedono la crescita di Pompei successivamente al trasferimento della nobiltà dalla capitale dell'impero, anche per fini militari, e dai ritrovamenti archeologici all'interno della città indicandola come una delle più attive città commerciali e portuali del bacino del Mediterraneo.

SUMMARY

The ancient city of Pompeii has a worldwide interest for its unique history, so that it is included in the UNESCO World Heritage. It represents a time capsule of Roman antiquity because buildings, monuments, sculptures, paintings and mosaics were preserved and remained intact, as well as food and personal objects related with daily life. Many human remains, belonging to people who died during the eruption of Vesuvius in 79 CE, have been unearthed thanks to the intense excavations of the archaeological site. Bone microstructures seem to be well preserved under 5-6 meters of pyroclastic materials, which also slowed down the DNA degradation process due to oxidative decay. Indeed, a first shallow shotgun sequencing of more than 170 human remains (mostly bones and in some cases even teeth) demonstrated a good preservation of ancient DNA (aDNA), exceeding even 90% in a few cases. This is the first time that a great number of individuals who lived or were in Pompeii for other reasons at the time of the eruption are studied from a population genomic perspective. In order to preserve the morphological structure of the archaeological human remains, we applied different protocols for sampling. In case of entire skull microsampling strategy was applied, while the auditory ossicles were selected instead of cutting the petrous bone when available. Ancient DNA was extracted using a specific silica-based protocol designed to recover short fragments of molecules. Double-stranded and double indexed genomic libraries were constructed with partial Uracil-DNA Glycosylase (UDG) treatment to repair spontaneous deaminations of ancient molecules, leaving only 2 base pairs (bp) damaged at both termini of the fragments as ancient signature. Next Generation Sequencing (NGS) platforms, such as Illumina MiSeq and NovaSeq machines, were used to sequence aDNA libraries.

Twelve genomic libraries with a good percentage of endogenous DNA preservation (>20%) were selected to perform a deep shotgun sequencing with the aim to obtain an average coverage of 1x. We used a whole-genome approach to analyse the genomic variability of individuals found in Pompeii and to make comparisons between them and other ancient Eurasian populations published so far. For the population genomic analyses, only the individuals with an average coverage equal or greater to 0.1x after shotgun sequencing and spanning a great temporal transect, from Mesolithic to Imperial age, were used. All specimens, including ancient populations data from literature, were processed in the same way using the same human reference genome (hg19) to avoid bias due to different processing pipeline. The Principal Component Analysis (PCA) demonstrates the presence of a heterogeneous populations. The individuals from Pompeii are localized not only on the genomic variability already found in Imperial populations from Italy, but also on the variability of England Roman settlements populations, France Imperial individuals and Italic populations during the Bronze and Iron age. These results are also confirmed by the Admixture analysis that shows a different distribution of the ancestral components along the genomes of the studied individuals from Pompeii. Unlike the previous periods, the "Italian Imperial" variability is featured by the absence of Steppe related ancestry and the increase

of Iranian Neolithic ones. In some outgroup cases these two different ancestral components are equally distributed, or Steppe related ancestry is predominant like in the genomes of the individuals of Bronze and Iron periods or in the coeval individuals of the foreign territories.

A parallel study looked at the human remains found in the same archaeological context and involved a multidisciplinary analysis on them. The anthropological study demonstrated the presence of human remains maybe belonged to 12 individuals, both adults and subadults, that found shelter in the so called "Stanza degli Scheletri" during the volcanic eruption. The forensic analysis on Short Tandem Repeats loci (STR) demonstrates the presence of 10 unique genomic profiles between the 12 human remains, so the minimum number of individuals in the room was ten. The genomic analyses were able to identify the sex at the molecular level and together with the anthropological findings it was possible to confirm the presence of adult women and subadults of both sexes. Moreover, we found a first-degree relationship between an adult and a child who sharing the same mitochondrial haplogroup could be mother daughter. Even in this small context we identified high genomic variability using Principal Component Analysis. The individuals are distributed along all major axes of modern variability and mother-daughter are also found to be located at distant points of the PCA, the first one near the Imperial French populations and the individuals from England Roman settlements, the second one near the classic Italian Imperial variability.

1 INTRODUZIONE

1.1 Pompei

Il Parco Archeologico di Pompei è scrigno di innumerevoli opere d'arte come monumenti, abitazioni affrescate, edifici pubblici, mosaici e sculture tanto da essere riconosciuto come Patrimonio dell'Umanità dall'UNESCO. Il dramma che ha interessato la città di Pompei ha fornito uno spaccato delle città della Roma Imperiale e un'occasione unica per conoscere la vita quotidiana delle civiltà romane; per questo può essere equiparata a una capsula del tempo.

Le origini di Pompei si fanno risalire all'VIII secolo a.e.c. quando gli Osci si insediarono alle pendici del Vesuvio e alla foce del fiume Sarno. In quell'epoca la città di Pompei aveva il ruolo di mediatrice tra differenti popolazioni quali quelle Greche, Etrusche e Sannitiche tanto da subire profonde influenze da quest'ultime, soprattutto dal punto di vista urbanistico. La città raggiunse però il massimo splendore tra il III e il II secolo a.e.c. poiché fu interessata da un ripopolamento che diede forte impulso alla costruzione di edifici pubblici e privati. In quel periodo venne anche eretto il tempio di Iside, chiara testimonianza dei contatti tra Pompei e l'Oriente. Infatti, in quello che fu denominato "periodo d'oro", aumentarono i traffici fluviali rendendo Pompei una tra le più importanti città scalo del Mediterraneo. Le favorevoli condizioni climatiche della zona e le caratteristiche del suolo consentirono inoltre la coltura dell'ulivo e della vite potendo esportare olio e vino fino in Provenza e Spagna. Inoltre, una delle più fiorenti attività della città consisteva nella trasformazione e lavorazione del grano (proveniente dalla Sicilia e dall'Africa) e altri cereali come il farro per la produzione del pane, come dimostrato dalla presenza di più di 34 forni molti dei quali dotati di mole asinarie per la molitura.

Pompei diventa però colonia Romana solo dopo l'assedio del pretore Lucio Cornelio Silla nell'89 a.e.c. nel corso delle Guerre Civili. In poco tempo la città si uniforma al contesto romano, sia dal punto di vista architettonico che istituzionale, divenendo la residenza di villeggiatura dei patrizi. In seguito, alcune famiglie favorevoli alla politica del primo imperatore romano Augusto si trasferirono a Pompei che, grazie a loro, vide la costruzione di numerosi edifici pubblici.

La città di Pompei, insieme a quella di Ercolano e le villeggiature di Oplontis e Stabia, venne distrutta dalla violentissima eruzione del complesso Somma-Vesuvio nel 79 e.c. che ne causò il sotterramento sotto uno spesso strato di cenere vulcanica e lapilli e successivamente alla quale si originò il Vesuvio, uno tra i più pericolosi vulcani attivi in Europa. Dalle testimonianze di Plinio il Giovane (Epistola, VI, 16; Epistola, VI, 20) sappiamo che l'eruzione, preceduta negli anni prima da forti terremoti, fu di tipo esplosivo e distinta in due fasi, la prima delle quali vide la rottura del cratere vulcanico con la risalita dei gas e della colonna piroclastica visibile anche da chilometri di distanza. Pompei fu subito interessata dalla pioggia di ceneri e lapilli che causarono anche il crollo di numerosi edifici all'interno dei quali molti abitanti si erano rifugiati. La seconda fase, durante la quale ogni essere ancora vivente perse la vita,

Inoltre, Fiorelli fu il primo che intuì la possibilità di ottenere calchi delle vittime dell'eruzione utilizzando del gesso liquido. Esso veniva colato all'interno della cenere indurita, lasciata vuota dai corpi decomposti, ottenendo dei calchi che riproducevano perfettamente l'aspetto, le ultime espressioni e posizioni di persone e animali rimasti a Pompei durante l'ultima fase dell'eruzione e investiti dal flusso piroclastico (*Figura 2*).



Figura 2 Corpi rinvenuti a Pompei di cui sono stati generati i calchi in gesso utilizzando la tecnica messa a punto dall'archeologo Giuseppe Fiorelli (immagini dal sito ufficiale del Parco Archeologico di Pompei).

1.1.1 Primi studi sul DNA antico dai resti umani di Pompei

Nel corso degli anni di scavo sono stati riportati alla luce più di mille corpi, rinvenuti sia negli strati di lapilli inferiori (probabilmente vittime del crollo degli edifici per peso dei materiali vulcanici depositati sui tetti), sia negli strati superiori (probabilmente le vittime delle ondate piroclastiche).

I resti umani e animali così ritrovati rappresentano una preziosa fonte di informazione anche sul piano genetico grazie alla possibilità di recupero del DNA antico (aDNA) dalle ossa e dai denti.

L'analisi del DNA nell'antica Pompei ha inizio con il ritrovamento di diversi corpi all'interno della Casa di Polibio (De Franciscis, 1988), il cui studio rientra all'interno di un progetto di ricerca multidisciplinare approvato dalle autorità archeologiche del Parco di Pompei che ha inizio nel 1995.

Le dinamiche dell'eruzione e le temperature a cui i corpi sono stati sottoposti non lasciavano presagire nulla di buono riguardo alla possibilità di recuperare materiale genetico dai resti umani. Ciononostante, essi presentavano una struttura ossea ben conservata grazie al pH leggermente basico della cenere vulcanica che si accumulò nel corso dell'eruzione. Le elevate temperature delle ondate piroclastiche causarono la disidratazione istantanea dei tessuti, ma le ossa rimasero conservate sotto uno spesso strato di lapilli (5-6 metri) che ne consentì la preservazione proteggendole dai diversi fattori ambientali e impedendo l'insorgenza di danni ossidativi nelle molecole di DNA (Matsuo et al., 1995).

Il primo studio in cui i resti ben conservati, sopra descritti, vennero analizzati a livello molecolare risale al 1998 (Cipollaro et al., 1998). Il lavoro degli antropologi fu coadiuvato dallo studio del DNA dimostrando la possibilità di amplificare la sequenza di un singolo gene, l'amelogenina, per la determinazione del sesso. Sebbene questa tecnica di analisi non sia più diffusa nel campo dell'aDNA, essa viene utilizzata, soprattutto in ambito forense, per stabilire il sesso di un individuo in quanto il gene che la codifica esiste sia sul cromosoma X (154 bp) sia sul cromosoma Y (106 bp). La presenza di una o due bande sul gel elettroforetico permette quindi di identificare rispettivamente una femmina e un maschio.

Rispetto al DNA nucleare, presente in una sola copia all'interno delle cellule, il DNA mitocondriale (mtDNA) non solo esiste in più copie all'interno dell'organulo, ma sono presenti anche più mitocondri all'interno delle cellule. Questa condizione, insieme alla caratteristica circolarità della molecola, consentono una migliore preservazione dell'mtDNA che viene quindi utilizzato negli studi di antropologia molecolare. Infatti, nel 2009 venne fornito un ulteriore contributo allo studio multidisciplinare sui resti umani ritrovati all'interno della Casa di Polibio proprio analizzando la regione ipervariabile (HVS1) dell'mtDNA (Di Bernardo et al., 2009), con lo scopo di identificare eventuali relazioni di parentela tra i 13 individui ritrovati al suo interno.

Solo nel 2022, grazie ai più fini metodi di estrazione dell'aDNA e alle nuove tecnologie di sequenziamento (vedi *paragrafo 1.3.1*), viene per la prima volta pubblicato il genoma 0.42x di un individuo ritrovato all'interno della "Casa del Fabbro" (Scorrano et al., 2022).

Ma sarà con i risultati di questo lavoro di tesi che per la prima volta verranno ampliate le conoscenze riguardo la genetica dell'antica popolazione di Pompei, o per meglio dire degli individui che, nei giorni dell'eruzione, si trovavano in città e non ebbero scampo.

1.2 Genetica delle popolazioni Europee

La variabilità genetica che caratterizza le attuali popolazioni Europee è il risultato di numerosi eventi migratori tra popolazioni diverse non solo culturalmente, ma soprattutto geneticamente. Le prime popolazioni di *H. sapiens* che si insediarono in Europa praticavano uno stile di vita basato sulla caccia e sulla raccolta, ma il loro contributo alle attuali popolazioni Europee risulta limitato a causa di una perdita di variabilità e soprattutto a causa dell'arrivo di nuove popolazioni (Marchi et al., 2022). Ciononostante, nelle moderne popolazioni Europee è possibile identificare un'*ancestry* legata ai cacciatori-raccoglitori (da qui in poi HGs dall'inglese *Hunter-Gatherers*) che risale a circa 37 mila anni fa (Seguin-Orlando et al., 2014). Da quel momento vi furono, nel continente, due principali transizioni culturali che ebbero forte ripercussione sul *pool* genico delle popolazioni: l'avvento dell'agricoltura durante il Neolitico e la rivoluzione dei metalli durante l'età del Bronzo (Günther and Jakobsson, 2016). Alle moderne popolazioni Europee è stato quindi fornito un contributo genetico soprattutto da tre distinte popolazioni ancestrali, ognuna delle quali ha contribuito in misura diversa anche in funzione della regione geografica analizzata (Lazaridis et al., 2014). Sebbene fosse inizialmente considerato un unico gruppo quello dei cacciatori-raccoglitori, essi mostrano una variabilità genetica che rispecchia la loro localizzazione geografica (Jones et al., 2015). Dopo l'*out of Africa* alcune popolazioni di HGs si spinsero fino in Europa originando un *cluster* abbastanza omogeneo indicato come WHG (cacciatori-raccoglitori dell'ovest), essi si separarono circa 45 mila anni fa dalle popolazioni di HGs che rimasero invece nella regione del Caucaso (CHG) e che contribuirono all'*ancestry* delle popolazioni di pastori nomadi delle Steppe Pontico-Caspiche. Dalla popolazione ancestrale che diede origine ai CHG si originò un ulteriore ramo rappresentate gli antenati dei primi agricoltori Neolitici del Levante (Jones et al., 2015). La rivoluzione Neolitica in Europa ebbe come risultato il passaggio da uno stile di vita dinamico basato sulla caccia e sulla raccolta del cibo a uno stile di vita sedentario legato allo sviluppo delle tecniche di agricoltura e allevamento per la produzione del cibo. Questa transizione fu spinta da una diffusione demica e non da un passaggio culturale; non fu quindi l'idea dell'agricoltura che arrivò in Europa dal Medio Oriente, ma furono gli agricoltori che spostandosi la introdussero (Ammerman and Cavalli-Sforza, 1984). A differenza delle popolazioni di HGs, gli agricoltori Neolitici sono geneticamente più vicini alle popolazioni dell'Europa sud-occidentale che non alle popolazioni del Vicino Oriente (Skoglund et al., 2012). Nel dettaglio, quest'affinità è particolarmente pronunciata con gli attuali Sardi e Baschi (Günther et al., 2015). È possibile identificare nei primi agricoltori dell'Anatolia e del Levante di 12 mila/11500 anni fa la popolazione sorgente dei gruppi Neolitici Europei (Mathieson et al., 2015; Omrak et al., 2016; Lazaridis et al., 2016; Kılınç et al., 2016).

Dati archeologici e genetici hanno suggerito che la diffusione dell'agricoltura in Europa avvenne tramite due rotte principali e di conseguenza con ritmi e modalità differenti. Una attraverso il Danubio (González-Fortes et al., 2017), l'altra attraverso la costa Mediterranea (Özdoğan, 2011). A

dimostrazione di una doppia diffusione vi fu anche il ritrovamento di utensili appartenenti a culture Neolitiche ben distinte: la cultura della ceramica lineare (LBK da *Linearbandkeramik*) in Ungheria ed Europa centrale e la cultura della ceramica impressa e cardiale nel Mediterraneo e nella penisola Iberica (Harris, 2017). Essendo più isolata dal continente Europeo, la Gran Bretagna vide l'arrivo delle tecniche di agricoltura e allevamento solo più tardi, circa 4000 anni a.e.c. (mille anni dopo il suo arrivo nelle regioni continentali) tramite il movimento di popoli discendenti dagli agricoltori Neolitici Iberici e quindi giunti nel continente attraverso la rotta Mediterranea (Brace et al., 2019).

Sebbene il flusso genico tra le popolazioni di agricoltori e quelle indigene di HGs fosse inizialmente limitato o addirittura assente (Jones et al., 2017), in alcune regioni, soprattutto nella fase finale del Neolitico, venne identificata la cosiddetta "rinascita Mesolitica" dovuta alla ricomparsa nei genomi dell'*ancestry* dei HGs, forse a causa della riduzione del flusso genico dall'Anatolia. L'adozione dell'allevamento e delle pratiche agricole da parte di queste popolazioni indigene portò gli studiosi ad accettare non solo un modello demico, ma anche un modello culturale di diffusione Neolitica (Zvelebil & Lillie, 2000; Fort, 2012; Fort, 2015). L'interazione tra i due popoli fu molto complessa soprattutto in Francia, punto d'incontro delle due rotte e nella quale venne identificato un mosaico di culture. Infatti, in questa regione, studi genetici hanno rilevato una maggiore presenza di *ancestry* dei HGs locali che risulta invece più diluita nelle regioni dell'Europa centro-orientale (Rivollat et al., 2020). A eccezione di alcune aree isolate, il *pool* genico degli Europei alla fine del Neolitico è una mescolanza di profili genici riconducibili agli allevatori-agricoltori Anatolici e in misura minore ai popoli di cacciatori-raccoglitori.

Gli arrivi dei HGs Mesolitici e degli agricoltori-allevatori Neolitici nel continente non furono gli unici episodi in cui lo spostamento massivo di popoli produsse un significativo cambio nel *pool* genico degli Europei. Infatti, a queste due principali componenti ancestrali, durante l'Età del Bronzo (BA dall'inglese *Bronze Age*), se ne aggiunse una terza proveniente dalle steppe dell'attuale Ucraina che contribuì a diversificare il profilo genetico dei popoli Europei nel corso della loro storia e ha avuto un peso maggiore nel nord che nel sud del continente (Haak et al., 2015), come si può osservare in *Figura 3*.

Testimonianze archeologiche hanno permesso di identificare nella popolazione di pastori semi-nomadi appartenenti alla cultura Yamnaya la popolazione che, dalle steppe Pontico-Caspiche intorno ai 4500 anni a.e.c., arrivò in Europa e permise lo sviluppo della cultura della Ceramica Cordata (Allentof et al., 2015; Haak et al., 2015; Narasimhan et al., 2019). In aggiunta all'espansione verso ovest, fu registrato un ulteriore spostamento a est come dimostrato dal ritrovamento, nella regione dei monti Altai e Saiani, di oggetti appartenenti alla cultura Afanasievo il cui profilo genetico della popolazione risulta indistinguibile da quello degli Yamnaya (Allentof et al., 2015). La diffusione in Europa di questi popoli fu probabilmente facilitata dall'impiego del carro a due ruote e dall'uso dei cavalli. Giunti nel continente introdussero una nuova componente genetica ancestrale denominata componente della steppa e fu a

partire dalla fine dell'Età del Bronzo che il *pool* genico degli attuali Europei cominciò a strutturarsi (Allentoft et al., 2015; Haak et al., 2015) (Figura 3).

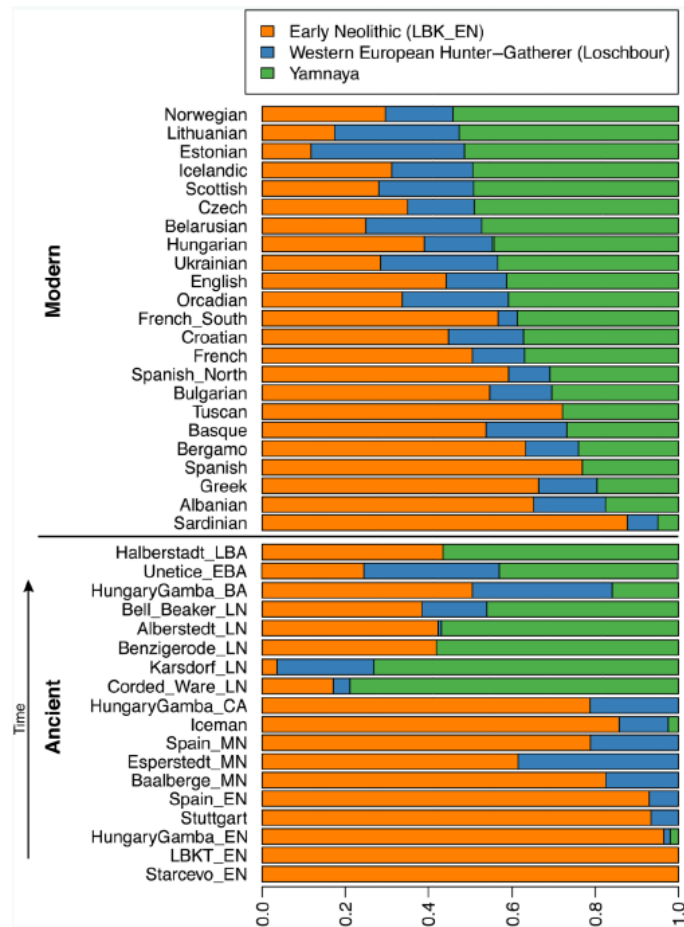


Figura 3 Proporzioe delle tre componenti principali nelle popolazioni antiche e moderne. A partire dall'Età del Bronzo compare nel genoma delle popolazioni Europee una nuova componente (in verde) legata alle migrazioni dalle Steppe. Da quel momento in Europa comincia a strutturarsi il *pool* genico delle popolazioni moderne. Il contributo della Steppa risulta maggiore nel nord del continente e diminuisce progressivamente nelle regioni del sud (immagine estratta da Haak et al., 2015).

Gli individui legati alla cultura Yamnaya portavano nel loro genoma delle componenti ben strutturate legate ai CHGs e ai cacciatori-raccoglitori dell'Est (EHGs) (Jones et al., 2015; De Barros Damgaard et al., 2018). In Europa quest'ulteriore componente genetica viene identificata in concomitanza con le migrazioni dalle Steppe. Gli individui appartenenti ai EHG mostrano affinità genetica con Mal'ta (MA1), il bambino Siberiano di 24 mila anni fa (Raghavan, Skoglund et al., 2014; Haak et al., 2015). Egli rappresenta l'olotipo per l'*Ancient North Eurasians* (ANE) i quali hanno contribuito alle attuali popolazioni Europee e ai Nativi Americani, ma non alle popolazioni Asiatiche (Raghavan, Skoglund et al., 2014).

Contemporaneamente alla presenza della cultura della Ceramica Cordata nell'Europa centro-settentrionale e nord-orientale, nella parte occidentale del continente, intorno ai 2750 anni a.e.c a

partire dalla penisola Iberica, si diffuse la cultura del Vaso Campaniforme (*Bell Beaker complex*). La diffusione di questa cultura nel resto del continente fu mediata sia da una trasmissione culturale, sia da flussi genici tra popoli vicini. In Iberia gli individui legati alla cultura *Bell Beaker* non mostrano nessuna *ancestry* della Steppa che è invece diffusa tra le popolazioni della stessa cultura nell'Europa centrale, suggerendo una diffusione di tipo culturale. Al contrario in Gran Bretagna l'arrivo della cultura *Bell Beaker* introdusse alti livelli della componente della Steppa, prima d'allora assente, determinando la sostituzione di circa il 90% del *pool* genico della popolazione (Olalde et al., 2018).

A partire dall'Età del Ferro (IA dall'inglese *Iron Age*) in poi, il *pool* genico delle popolazioni Europee comincia a diventare simile a quello ritrovato oggi nelle popolazioni delle stesse regioni (Günther and Jakobsson, 2016; Antonio et al., 2023 bioRxiv; Moots et al., 2023). Tra di esse non vi era una struttura genetica come quella presente tra le popolazioni Mesolitiche di HGs o come quelle di agricoltori Neolitici (Günther and Jakobsson, 2016). Infatti, durante il periodo storico (a partire da 3 mila anni a.e.c) si osserva nel genoma delle popolazioni del continente alta eterogeneità genica dovuta a una maggiore mobilità (Antonio et al., 2019; Olalde et al., 2019). Questo è in linea con i dati archeologici e storici i quali dimostrano, durante l'Età del Ferro, un massivo utilizzo delle navi per il commercio nel Mediterraneo (Abulafia, 2011; Broodbank, 2013). I movimenti di uomini aumentarono ancora di più durante il periodo della Roma Imperiale la cui organizzazione militare, commerciale e lavorativa richiamò persone sia dall'interno che dall'esterno dell'impero. Nonostante l'intensa mobilità, la composizione genetica delle popolazioni del continente rimase relativamente stabile nel tempo (Antonio et al., 2023 bioRxiv).

1.2.1 L'Italia Imperiale dal punto di vista genomico

Il primo ampio transetto temporale (che copre gli ultimi 12 mila anni) in cui viene studiato con approccio *shotgun* un consistente numero di resti umani ritrovati a Roma e dintorni porta la firma di Antonio e collaboratori (2019).

Come descritto nel paragrafo precedente, il periodo in cui si diffonde l'uso dei metalli nel continente Europeo e in Italia vede profondi cambiamenti non solo nella cultura delle popolazioni, ma anche nel loro *pool* genico. Questo cambiamento è evidente principalmente nel passaggio dall'Età del Rame (CA dall'inglese *Copper Age*) all'Età del Bronzo come risultato dell'introduzione dell'*ancestry* della Steppa, intorno a 1650 anni a.e.c, come dimostrato per la prima volta dallo studio genomico effettuato su individui ritrovati in territorio italiano del BA (Saupe et al., 2021). Nel successivo periodo dell'IA la popolazione del medesimo territorio mantiene la stessa struttura genetica dell'epoca precedente (*Figura 4*).

Nel periodo Imperiale, epoca in cui si collocano gli ultimi abitanti di Pompei, si osserva non solo una grande variabilità genetica tra gli individui, ma un'altra grande trasformazione del *pool* genico della

popolazione del territorio italiano. Infatti, analizzando la PCA (Analisi delle Componenti Principali) in *Figura 4*, gli individui di epoca Imperiale (in azzurro) si sovrappongono alle attuali popolazioni del Mediterraneo centro-orientale come conseguenza dell'aumento nel loro genoma dell'*Iran Neolithic ancestry*, componente massimizzata nei primi gruppi Neolitici di pastori dell'area iraniana e identificata in Sicilia già nel II millennio a.e.c. (Fernandes et al., 2020).

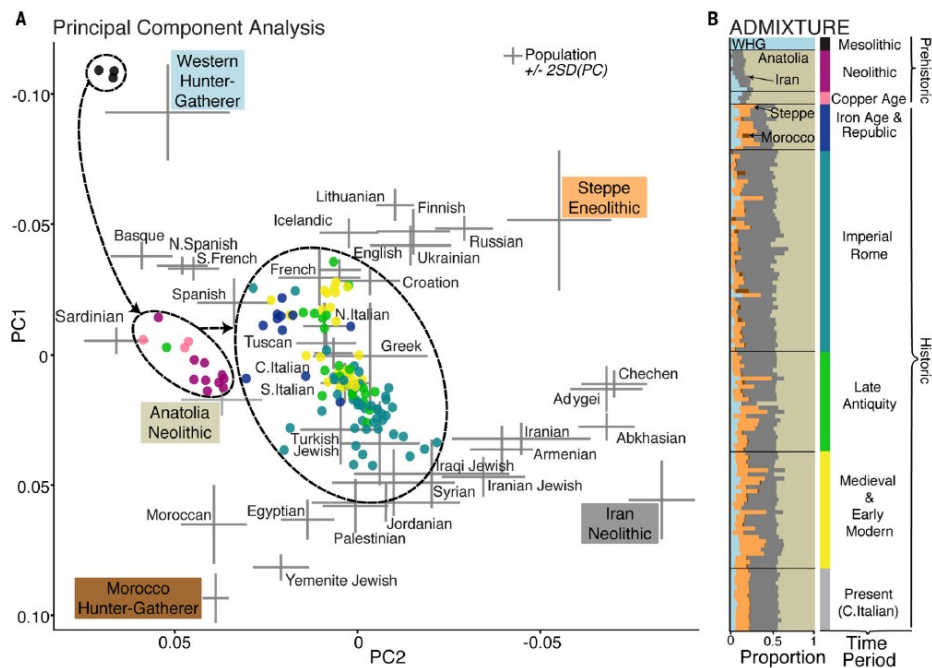


Figura 4 Analisi delle componenti principali (PCA) e di Admixture su individui italiani di un ampio transetto temporale. In PCA gli individui del CA sono sovrapposti alle attuali popolazioni Sarde e il loro pool genico riflette il precedente periodo Neolitico. In PCA l'arrivo della componente della Steppa (marcata in arancione nell'analisi di Admixture) si riflette in uno *shift* degli individui (in blu) che si collocano ora sulle attuali popolazioni dell'Europa dell'ovest (sebbene non ci siano individui del BA questa considerazione può essere comunque effettuata alla luce del successivo lavoro di Saupé et al., 2021 e sul mantenimento della stessa struttura genetica nell'IA) (immagine estratta da Antonio et al., 2019).

1.3 Il DNA antico

Il genoma umano, dalla comparsa di *H. sapiens* 200 mila anni fa (forse anche 300 mila) in Africa fino ai giorni d'oggi, tiene traccia dei numerosi e complessi eventi che si sono susseguiti nel tempo e che hanno permesso di generare la grande variabilità genetica della nostra specie. La possibilità di individuare e associare tali segni a momenti chiave nel corso della preistoria e della storia è affidata allo studio del DNA antico (aDNA).

Con aDNA si indicano le molecole di acido desossiribonucleico presenti in resti umani, animali o vegetali che, successivamente alla morte cellulare, vanno incontro a un processo di degradazione. I grandi passi in avanti nelle tecnologie di estrazione dell'aDNA hanno fatto sì che questo si potesse recuperare anche partendo da materiale esterno all'organismo come il suolo, i coproliti, o le resine (Willerslev et al., 2003; Gilbert et al., 2008; Jensen et al., 2019).

L'era degli studi sull'aDNA ha inizio nel 1984 quando il gruppo diretto da Allan Wilson, in California, riesce a estrarre e leggere corte sequenze di DNA mitocondriale (229 bp) da un esemplare estinto di equide vicino alla zebra, l'*Equus quagga* (Higuchi et al., 1984). Contemporaneamente in Svezia Svante Pääbo, premio Nobel per la Medicina nel 2022, riesce a isolare il DNA da tessuti umani mummificati e antichi più di 2 mila anni (Pääbo, 1985). Col passare del tempo è stato possibile ricostruire il genoma di numerose specie e popolazioni estinte, come i mammut lanosi, gli orsi delle caverne, l'uomo di Neanderthal e di Denisova (Palkopoulou et al., 2015; Barlow et al., 2018; Green et al., 2010; Prüfer et al., 2014; Reich et al., 2010; Meyer et al., 2012). L'aDNA è quindi uno strumento di fondamentale importanza per individuare eventi che, con il solo studio dei reperti archeologici, non potrebbero essere ricostruiti con grande dettaglio.

Lo studio dell'aDNA presenta però numerose difficoltà legate alla sua stessa natura e alla sua conservazione nel tempo. Durante la replicazione del DNA può capitare che vengano introdotti errori casuali che, se non riparati, si fissano nel genoma. Il processo di mutazione può essere scatenato anche da agenti esterni, chimici e fisici. Una cellula viva possiede numerosi meccanismi di riparazione che fanno sì che gli errori vengano individuati e corretti, ma dopo la morte cellulare essi vengono inattivati e il DNA va incontro a processi diagenetici, primo dei quali l'azione delle nucleasi (degradazione autolitica). La degradazione delle molecole di DNA è velocizzata in ambienti caldi e umidi a pH acido e/o basico, mentre in zone più fredde a pH neutro e in assenza di ossigeno e acqua i processi degradativi risultano rallentati (Hofreiter et al., 2015). Questi processi determinano così la comparsa dei pattern di degrado che caratterizzano le molecole di aDNA (Dabney et al., 2013) i quali, sebbene in misura variabile, risultano correlati all'età del campione (Sawyer et al., 2012).

L'aDNA si presenta frammentato e la principale causa sembra essere la depurinazione che porta alla rimozione della base azotata per taglio del legame N-glicosidico. La rottura dello scheletro zuccherofosfato è poi dovuta alla β eliminazione che lascia libere le estremità 3'-aldeidico e 5'-fosfato (*Figura*

5). L'elaborazione di nuovi protocolli nella costruzione delle librerie ha infatti dimostrato che la rottura della molecola avviene preferenzialmente prima e dopo le guanine (Meyer et al., 2012).

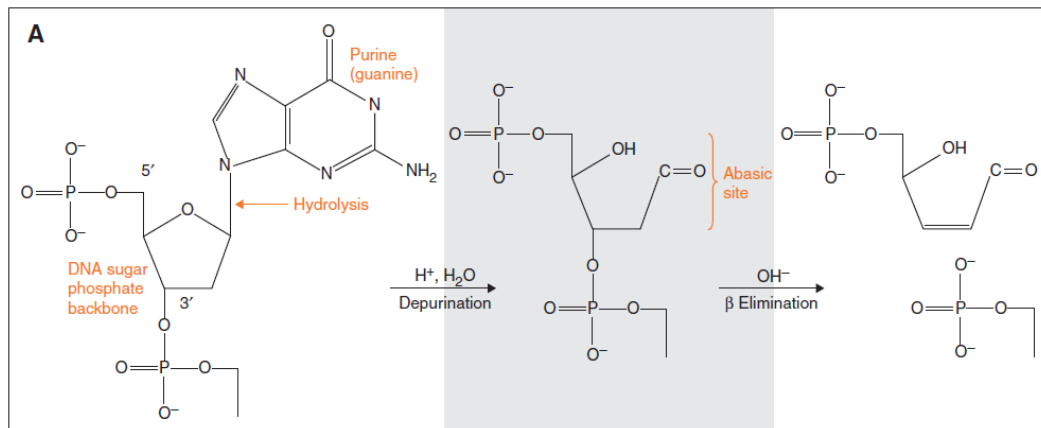


Figura 5 Idrolisi del legame N-glicosidico tra lo zucchero e la base azotata con formazione del sito abasico e β eliminazione con rottura dell'impalcatura zucchero-fosfato (immagine estratta da Dabney et al., 2013).

Proprio in prossimità delle estremità dei frammenti si verifica, con maggiore frequenza, il processo di deamminazione che coinvolge le citosine. Questa è una conseguenza della maggiore esposizione delle estremità dei frammenti, spesso a singolo filamento, ai processi diagenetici (Briggs et al., 2007). L'idrolisi del gruppo amminico causa la conversione della citosina in uracile (Hofreiter et al., 2001) (Figura 6). La presenza della U provoca, in fase di sequenziamento, l'incorporazione dell'adenina con conseguente sostituzione C->T G->A (sul filamento complementare) (Brotherton et al., 2007).

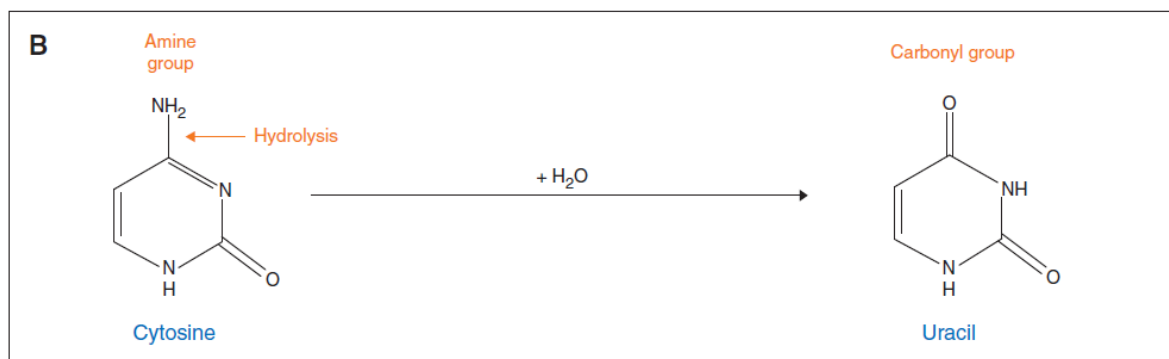


Figura 6 Le citosine subiscono, in presenza di acqua, una spontanea deamminazione con conseguente trasformazione in uracile (immagine estratta da Dabney et al., 2013).

Come precedentemente accennato, l'adna subisce anche modifiche dovute all'azione di agenti esterni come le radiazioni ultraviolette (UV) le quali provocano la fusione dei nucleotidi, anche tra filamenti

diversi di DNA o tra DNA e altre molecole. La conseguenza è la formazione di dimeri che impediscono alla DNA polimerasi di scorrere lungo i frammenti durante le fasi di amplificazione e sequenziamento. Sebbene sia stata trovata una correlazione tra l'età del campione e la quantità di purine alle estremità dei frammenti di aDNA, nessuna connessione è stata invece trovata tra l'età e la lunghezza dei frammenti (Dabney et al., 2013). Solo il tasso di deaminazione delle citosine è positivamente correlato all'età degli individui, nonostante i resti possano trovarsi in ambienti a condizioni climatiche differenti. La frequenza di sostituzioni C-> T può quindi essere utilizzata come indicatore dell'autenticità della molecola antica (Krause et al., 2010).

Proprio a causa dei processi diagenetici la conservazione del DNA nel tempo è ostacolata, infatti, la quantità di DNA endogeno (proprio del campione analizzato) che si riesce a recuperare è scarsa e in alcuni casi nulla. Proprio per ottimizzare il recupero dell'aDNA non solo sono stati sviluppati protocolli per l'estrazione delle corte sequenze di DNA endogeno (Dabney et al., 2013), ma sono anche stati individuati i distretti ossei e gli elementi migliori per ottenere una maggiore resa in termini di quantità di DNA umano recuperato. L'elemento più proficuo da cui estrarre il DNA e da cui ottenere la percentuale più alta di molecole è stato individuato nella rocca petrosa (Pinhasi et al., 2015) (Figura 7) – la quale dona una resa dalle 4 alle 16 volte maggiore rispetto a quella ottenuta dai denti e 183 volte maggiore rispetto a quella ottenuta da altri tipi di ossa (Gamba et al., 2014) – localizzata internamente all'osso temporale.

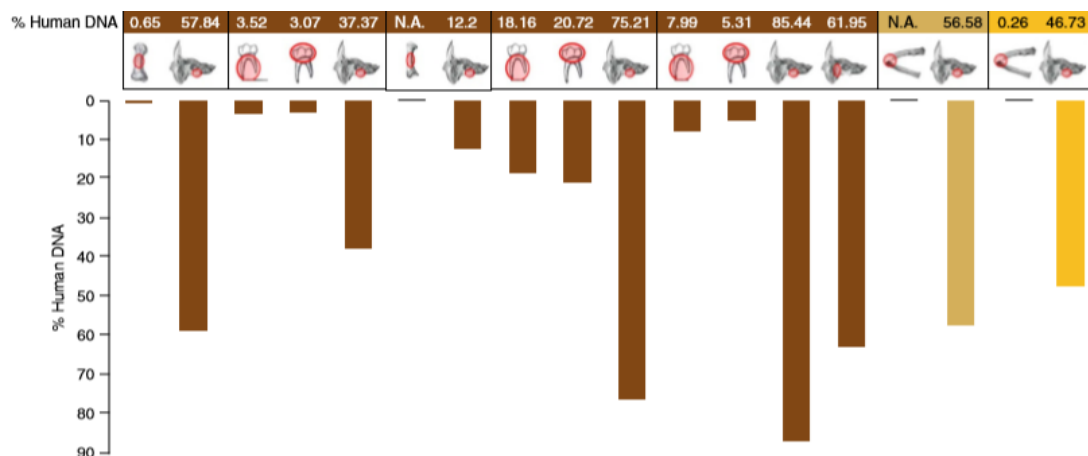


Figura 7 La quantità di DNA endogeno ricavata dalla rocca petrosa varia da 37.4% a 85.4% mentre nei denti o in altri resti post-craniali il DNA è scarso o assente (immagine estratta da Gamba et al., 2014).

Nel cuore della rocca petrosa è ubicato l'orecchio interno comprendente il sistema vestibolare (canali semicircolari e vestibolo) e la coclea, porzioni dalle quali si recupera una maggiore quantità di DNA endogeno (Pinhasi et al., 2015). Procedendo verso l'interno a partire dal meato acustico (base della rocca) si trova il cavo timpanico contenente gli ossicini dell'udito (martello, incudine e staffa) i quali

sono stati proposti come un'ottimale e alternativa fonte di estrazione dell'aDNA; infatti, da essi è stata dimostrata la possibilità di recuperare una percentuale maggiore di molecole endogene rispetto a quella ottenuta dalla coclea polverizzata (Sirak et al., 2022).

L'elevata degradazione dell'aDNA, la sua scarsa preservazione e la manipolazione dell'elemento da parte di diversi soggetti potenzialmente contaminanti (DNA moderno), impongono l'adozione di specifiche accortezze durante le fasi di analisi sperimentale. Per ridurre al minimo il rischio di contaminazione sono state stilate nei *Golden Criteria* (Cooper & Poinar, 2000) le norme comportamentali da tenere all'interno di un laboratorio in cui si manipola aDNA. Queste prevedono regole sul vestiario e sul comportamento degli operatori, la suddivisione fisica dei laboratori in spazi adibiti alle diverse fasi di analisi (aree a contaminazione controllata pre e post amplificazione) e l'utilizzo di controlli negativi nelle fasi di estrazione del DNA e preparazione delle librerie. L'ingresso nelle aree di lavoro è consentito solo indossando tute di protezione complete, guanti, mascherine e calzari. Le superfici all'interno dei laboratori vengono mantenute in condizioni di sterilità per esposizione giornaliera alle radiazioni ultraviolette (UV) e trattamento con candeggina. Gli strumenti di laboratorio vengono sterilizzati, prima e dopo il loro utilizzo, con etanolo e candeggina, mentre tutti gli esperimenti vengono svolti sotto cappe a flusso laminare continuo con sistemi di aspirazione (in caso di produzione delle polveri) e filtraggio dell'aria per impedire contaminazioni. Inoltre, risulta altrettanto importante tracciare eventuali episodi di contaminazione avvenuti fuori dal laboratorio, come nei siti di scavo e di conservazione dei resti oggetto delle analisi (Pilli et al., 2013).

1.3.1 Tecniche di analisi dell'aDNA

Un forte contributo allo studio dell'aDNA è stato fornito dall'utilizzo della tecnica della PCR (*Polymerase Chain Reaction*), ideata da Kary Mullis nel 1986. La PCR permette l'amplificazione esponenziale di sequenze target di DNA e proprio grazie ad essa è stato possibile aumentare la scarsa concentrazione delle molecole, provenienti da resti antichi, raggiungendo quantità necessarie per le successive analisi (Pääbo, 1989). L'uso della PCR era però limitato da diversi fattori, primo fra tutti l'utilizzo di *primers* (20-30 bp) specifici per l'amplificazione di sequenze normalmente troppo lunghe (~150 bp) per le dimensioni dei frammenti di aDNA. Inoltre, per poter disegnare i primer, è necessario conoscere in partenza parte della sequenza da amplificare, cosa che non è garantita in campioni antichi, che possono presentare varianti genetiche assenti in genomi contemporanei. Infine, la presenza di danni molecolari caratteristici delle molecole antiche favoriva l'amplificazione preferenziale delle molecole di DNA moderno più lunghe e meno danneggiate (Willerslev & Cooper 2005; Knapp e Hofreiter 2010). Ma è stato l'avvento delle NGS (*Next Generation Sequencing*) agli inizi degli anni 2000 a trasformare completamente lo studio dell'aDNA. Il vantaggio fornito dalle tecniche di sequenziamento di nuova generazione è stato la possibilità di leggere contemporaneamente centinaia di milioni di corte

sequenze (35÷300 bp) di DNA, senza l'utilizzo di *primers*, tramite l'impiego di adattatori legati alle estremità dei frammenti e complementari a degli oligonucleotidi presenti nella *flowcell* della macchina Illumina.

In pochi anni si è quindi passati dallo studio di corte sequenze di DNA mitocondriale (più abbondante e quindi facilmente recuperabile) a studi di intere popolazioni antiche su scala genomica (Willerslev & Cooper 2005).

Le strategie di sequenziamento delle librerie sono però molteplici e dipendono dalla preservazione del DNA nel campione, dalle risorse economiche messe in campo e dall'obiettivo dello studio.

In ogni caso, si procede sempre con un sequenziamento *shotgun* esplorativo per valutare la preservazione del DNA all'interno del campione in esame. 5 milioni di *reads* generate per campione sono sufficienti per raggiungere tale scopo e per valutare le misincorporazioni tipiche delle molecole antiche.

Quando la percentuale di DNA endogeno risulta elevata (>20%) è possibile realizzare un sequenziamento di tipo *shotgun* profondo, senza previa applicazione di metodi di arricchimento, il quale consente la lettura di tutte le molecole di DNA rappresentando la strategia migliore per un'analisi completa (sia a livello nucleare che mitocondriale). In alternativa, la strategia del *target enrichment*, la quale prevede l'utilizzo di sonde oligonucleotidiche, consente di operare tramite un arricchimento selettivo di determinate regioni genomiche producendo non solo dati su singoli loci, ma anche a livello cromosomico e genomico, sia esso nucleare o mitocondriale (Ávila-Arcos et al., 2011; Fu et al., 2013; Carpenter et al., 2013; Maricic et al., 2010). Alcuni dei vantaggi delle tecniche di cattura consistono nell'abbassare i costi di sequenziamento rispetto alla strategia *shotgun* (la quale richiede una profondità di sequenziamento elevata) e nella possibilità di analizzare anche campioni con bassa percentuale di DNA endogeno ($\geq 0.15\%$) che altrimenti verrebbero scartati. Inoltre, con quest'ultima tecnica è possibile ridurre la produzione di sequenze di DNA contaminante o appartenente a organismi che non rappresentano il target dello studio. Con il *target enrichment* però risulta impossibile individuare polimorfismi diversi da quelli già noti in partenza e quindi rappresentati dalle sonde usate per la cattura.

2 OBIETTIVO

L'antica città di Pompei ha sempre suscitato un immenso fascino nel mondo accademico e non solo proprio per la sua storia unica al mondo.

Essere rimasta immobile, ferma a un attimo preciso della storia, ha fatto in modo che potessero essere studiate, tramite dei riscontri diretti rimasti lì immutati nel tempo, le abitudini (comprese quelle alimentari), lo stile di vita (lussuoso per le figure di rilievo della città e semplice per il popolo), la cultura, ma non solo. Lo studio del DNA delle vittime dell'eruzione ha suscitato l'interesse dei ricercatori fin dai primi anni '90 (come già discusso nel *paragrafo 1.1.1*). È solo nello studio del 2022 (Scorrano et al., 2022) che per la prima volta viene effettuato uno studio genomico su scala popolazionistica, limitato però a una sola vittima dell'eruzione del Vesuvio. Nonostante gli sforzi messi in atto per definire un "portrait" molecolare degli individui che, nel periodo immediatamente precedente all'eruzione, conducevano la propria vita a Pompei, si è reso necessario aumentare il numero di campioni da analizzare affinché ciò che emerge abbia un maggiore supporto statistico.

Questo studio nasce grazie alla convenzione tra il dipartimento di Biologia dell'Università degli Studi di Firenze e il Laboratorio di Ricerche Applicate del Parco Archeologico di Pompei e ha visto, tra il 2019 e il 2022, il campionamento di numerosi resti umani destinati allo studio dell'adDNA. La possibilità di processare un numero elevato di campioni, provenienti dallo stesso sito, ha fatto sì che le analisi genomiche potessero essere condotte sui soli individui caratterizzati da una migliore e maggiore preservazione delle molecole di DNA endogene per raggiungere una migliore risoluzione delle analisi. Il primo obiettivo di questo studio è stato quindi quello di realizzare un campionamento mirato ad ottenere una buona polvere (da ossa o denti) preservando, quando possibile, l'integrità del reperto e applicare sulle librerie genomiche uno studio di ultima generazione tramite le tecniche di sequenziamento NGS. Questo lavoro è stato realizzato sia su individui ritrovati in vari luoghi della città, sia su individui facente parte di particolari e meglio definiti contesti di scavo.

Il secondo obiettivo specifico dello studio è stato quello di realizzare sequenziamenti a profondità differenti a seconda del caso studio esaminato.

In seguito, la ricerca ha visto il suo prosieguo nello sfruttamento di approcci genomici e antropologici per la risoluzione di particolari dinamiche che hanno visto il formarsi di piccoli gruppi di persone, probabilmente unite da legami affettivi di varia natura come quelli di parentela, in luoghi riparati della città.

Infine, si è proceduto con la selezione di alcuni degli individui ritrovati nelle varie aree della città (probabilmente persone in via di fuga) per condurre analisi popolazionistiche utilizzando un approccio *whole-genome*, compatibilmente con la preservazione delle molecole che comunque portano i segnali dell'adDNA.

Il fine ultimo del lavoro di tesi è stato quello di dimostrare e mettere in risalto la diversità non solo culturale, ma anche genomica degli abitanti dell'antica città di Pompei la quale, come anche attestato dai ritrovamenti archeologici e dai documenti rinvenuti, rappresentava una delle più importanti città portuali del Mediterraneo.

3 MATERIALI E METODI

3.1 Materiale Archeologico

In seguito alla convenzione fra il Parco Archeologico di Pompei e il Dipartimento di Biologia dell'Università degli Studi di Firenze, sono stati processati 173 resti umani provenienti dal Laboratorio di Ricerche Applicate di Pompei risalenti a campagne di scavo collocate in epoche temporali diverse e localizzate in differenti zone dell'antica città. Sono stati trattati i soli resti degli individui che persero la vita nel corso dell'eruzione del 79 e.c. e sono state applicate diverse procedure di campionamento in funzione del reperto in esame. Le informazioni riguardanti il distretto osseo analizzato, l'ubicazione all'interno del Parco Archeologico e l'ID del campione sono riportate in *Tabella 1*.

ID Archeologico	ID Laboratorio	Elemento scheletrico	Sito Archeologico
US 11A Nicchia	Pom-1	Cranio	Stanza degli Scheletri
US416 (A1)	Pom-2	Cranio	Stanza degli Scheletri
II7	Pom-3	Rocca petrosa dx	Stanza degli Scheletri
III 1	Pom-4	Rocca petrosa sx	Stanza degli Scheletri
11A	Pom-5	Rocca petrosa sx	Stanza degli Scheletri
11A 1	Pom-6	2° molare inf. sx deciduo	Stanza degli Scheletri
364/2	Pom-7	1° molare inf. sx deciduo	Stanza degli Scheletri
II 6 = 416	Pom-8	1° premolare sup. sx	Stanza degli Scheletri
II6	Pom-8.1	Rocca petrosa sx	Stanza degli Scheletri
II6	Pom-8.2	1° incisivo sup. sx	Stanza degli Scheletri
US 356	Pom-9	1° molare inf. sx deciduo	Stanza degli Scheletri
III1 (A2)	Pom-10	1° premolare inf. dx	Stanza degli Scheletri
III1 II8	Pom-10.1	1° premolare inf. sx	Stanza degli Scheletri
III1 II7	Pom-10.1	Rocca petrosa dx	Stanza degli Scheletri
III2	Pom-11	2° incisivo inf. dx	Stanza degli Scheletri
III2	Pom-11.1	1° incisivo inf. dx	Stanza degli Scheletri
III7 (A3)	Pom-12	Rocca petrosa sx	Stanza degli Scheletri
GPPM vittima 1 (US 1111)	Pom-13	Cranio	Isolato (GPPM)
Individuo 1 Terme centrali	Pom-14	Rocca petrosa sx	Terme centrali
Tomba 1 da Civita Giuliana	Pom-15	Rocca petrosa dx	Tomba 1 Civita Giuliana
GPPM US 442	Pom-16A	Rocca petrosa sx	Tunnel vicino Stanza Scheletri (GPPM)
GPPM 442	Pom-16B	3° molare inf. Sx	Tunnel vicino Stanza Scheletri (GPPM)
Porta Capua Ind.1	Pom-17	Rocca petrosa	Porta Capua
Porta Capua Ind.2	Pom-18	Rocca petrosa	Porta Capua
GPPM Fronte V, 7 US 54 (10/08/18)	Pom-19	Rocca petrosa	Scavi nuovi, interferenza Borbonica
GPPM US 312 Inv. PM 2080	Pom-20	Rocca petrosa	Scavi nuovi, interferenza Borbonica
GPPM R V, 7 US 54 amb 7 (07/05/2018)	Pom-21	Rocca petrosa	Scavi nuovi, interferenza Borbonica
GPPM Stanza scheletri	Pom-22	Osso lungo	Stanza degli scheletri
L16/03	Pom-23	Rocca petrosa	Terme del Sarno
L22/03	Pom-24	Rocca petrosa	Terme del Sarno
L35/02	Pom-25	Rocca petrosa	Terme del Sarno
L86/02	Pom-26	Rocca petrosa	Terme del Sarno
L67/02	Pom-27	Rocca petrosa	Terme del Sarno
L08/03	Pom-28	Rocca petrosa	Terme del Sarno
L21/03	Pom-29	Rocca petrosa	Terme del Sarno
L23/03	Pom-30	Rocca petrosa	Terme del Sarno
L32/03	Pom-31	Rocca petrosa	Terme del Sarno
L02/03	Pom-32	Rocca petrosa	Terme del Sarno
L15/02	Pom-33	Rocca petrosa	Terme del Sarno

ID Archeologico	ID Laboratorio	Elemento scheletrico	Sito Archeologico
L18	Pom-34	Rocca petrosa	Terme del Sarno
L104/03	Pom-35	Rocca petrosa	Terme del Sarno
L11/02	Pom-36	Rocca petrosa	Terme del Sarno
L25/03	Pom-37	Rocca petrosa	Terme del Sarno
L56/02	Pom-38	Rocca petrosa	Terme del Sarno
L33/02	Pom-39	Rocca petrosa	Terme del Sarno
L14/03	Pom-40	Rocca petrosa	Terme del Sarno
L69/02	Pom-41	Rocca petrosa	Terme del Sarno
L31/02 (198/599)	Pom-42	Rocca petrosa	Terme del Sarno
L21/02	Pom-43	Rocca petrosa	Terme del Sarno
L61/02	Pom-44	Rocca petrosa	Terme del Sarno
L19/03	Pom-45	Rocca petrosa	Terme del Sarno
L68/02	Pom-46	Rocca petrosa	Terme del Sarno
L54/02	Pom-47	Rocca petrosa	Terme del Sarno
L53/02	Pom-48	Rocca petrosa	Terme del Sarno
R25/02	Pom-49	Rocca petrosa	Terme del Sarno
L01/03	Pom-50	Rocca petrosa	Terme del Sarno
L10/03	Pom-51	Rocca petrosa	Terme del Sarno
L85/02	Pom-52	Rocca petrosa	Terme del Sarno
L46/03	Pom-53	Rocca petrosa	Terme del Sarno
L24/03	Pom-54	Rocca petrosa	Terme del Sarno
L18/03	Pom-55	Rocca petrosa	Terme del Sarno
L09/03	Pom-56	Rocca petrosa	Terme del Sarno
R24/02	Pom-57	Rocca petrosa	Terme del Sarno
L96/02	Pom-58	Rocca petrosa	Terme del Sarno
RA	Pom-59	Rocca petrosa	Terme del Sarno
TDS 135	Pom-60	Rocca petrosa	Terme del Sarno
L36/02	Pom-61	Rocca petrosa	Terme del Sarno
L92/02	Pom-62	Rocca petrosa	Terme del Sarno
L12/08	Pom-63	Rocca petrosa	Terme del Sarno
L19/02	Pom-64	Rocca petrosa	Terme del Sarno
L01/20	Pom-65	Rocca petrosa	Terme del Sarno
L95/02	Pom-66	Rocca petrosa	Terme del Sarno
L44/03	Pom-67	Rocca petrosa	Terme del Sarno
L36/02	Pom-68	Rocca petrosa	Terme del Sarno
L91/02	Pom-69	Rocca petrosa	Terme del Sarno
L5/02	Pom-70	Rocca petrosa	Terme del Sarno
SB10/78	Pom-71	Rocca petrosa	Terme del Sarno
L05/03	Pom-72	Rocca petrosa	Terme del Sarno
L06/03	Pom-73	Rocca petrosa	Terme del Sarno
L11/02	Pom-74	Rocca petrosa	Terme del Sarno
L27/02	Pom-75	Rocca petrosa	Terme del Sarno
L12/01	Pom-76	Rocca petrosa	Terme del Sarno
L15/03	Pom-77	Rocca petrosa	Terme del Sarno
L65/02	Pom-78	Rocca petrosa	Terme del Sarno
L47/03	Pom-79	Rocca petrosa	Terme del Sarno
L37/02	Pom-80	Rocca petrosa	Terme del Sarno
L11/03	Pom-81	Rocca petrosa	Terme del Sarno
L22/02	Pom-82	Rocca petrosa	Terme del Sarno
L14/02	Pom-83	Rocca petrosa	Terme del Sarno
L40/02	Pom-84	Rocca petrosa	Terme del Sarno
LR82/02	Pom-85	Rocca petrosa	Terme del Sarno
L49/02	Pom-86	Rocca petrosa	Terme del Sarno
L89/02	Pom-87	Rocca petrosa	Terme del Sarno
L07/03	Pom-88	Rocca petrosa	Terme del Sarno
104	Pom-89	Rocca petrosa	Terme del Sarno

ID Archeologico	ID Laboratorio	Elemento scheletrico	Sito Archeologico
SIM.3 casti amanti	Pom-90	Cranio	Casti amanti
SCH.3 casti amanti	Pom-91	Molare incisivo	Casti amanti
SCH.1 casti amanti	Pom-92	Osso piatto	Casti amanti
SCH.2 casti amanti	Pom-93	Rocca petrosa	Casti amanti
Termopolio Ind. 1	Pom-95B	1° premolare inf. sx	Termopolio
Termopolio Ind. 2	Pom-96A	2° premolare sx	Termopolio
Termopolio Ind. 2	Pom-96B	Rocca petrosa	Termopolio
US170 Ind. 1	Pom-97	Dente	Civita Giuliana
US173 Ind. 2	Pom-98	Falange	Civita Giuliana
ID 99 MLF1	Pom-99	Cranio	Terme Femminili
ID 100 MLF2	Pom-100	Cranio	Terme Femminili
ID 101 TF31	Pom-101	Cranio	Terme Femminili
ID 102 108, 16, TF60	Pom-102	Cranio	Terme Femminili
ID 103 26, TF107	Pom-103	Cranio	Terme Femminili
ID 104 07, TF86	Pom-104	Cranio	Terme Femminili
ID 105 57, TF17	Pom-105	Cranio	Terme Femminili
ID 106 TF85	Pom-106	Cranio	Terme Femminili
ID 107 TF95	Pom-107	Cranio	Terme Femminili
ID 108 TF82	Pom-108	Cranio	Terme Femminili
ID 109 TF61	Pom-109	Cranio	Terme Femminili
ID 110 97, TF4	Pom-110	Cranio	Terme Femminili
ID 111 TF91	Pom-111	Cranio	Terme Femminili
ID 112 22, TF22	Pom-112	Cranio	Terme Femminili
ID 113 91, TF98	Pom-113	Cranio	Terme Femminili
ID 114 18	Pom-114	Cranio	Terme Femminili
ID 115 80	Pom-115	Cranio	Terme Femminili
ID 116 112	Pom-116	Cranio	Terme Femminili
ID 117 05, TF39	Pom-117	Cranio	Terme Femminili
ID 118	Pom-118	Cranio	Terme Femminili
ID 119 TF89	Pom-119	Cranio	Terme Femminili
ID 120 TF106	Pom-120	Cranio	Terme Femminili
ID 121 TF100	Pom-121	Cranio	Terme Femminili
ID 122 54, TF115	Pom-122	Cranio	Terme Femminili
ID 123 03,TF112	Pom-123	Cranio	Terme Femminili
ID 124 TF96	Pom-124	Cranio	Terme Femminili
ID 125 TF83	Pom-125	Cranio	Terme Femminili
ID 126 124, TF87	Pom-126	Cranio	Terme Femminili
ID 01 72, 160	Pom-127	Cranio	Terme Femminili (?)
ID 02 (?) TF47	Pom-128	Cranio	Terme Femminili (?)
ID 03 30, TF7	Pom-129	Cranio	Terme Femminili (?)
ID 05 TF29	Pom-130	Cranio	Terme Femminili (?)
ID 06 53,74, TF13	Pom-131	Cranio	Terme Femminili (?)
ID 07 TF34	Pom-132	Cranio	Terme Femminili (?)
ID 08 19, IND NO 14	Pom-133	Cranio	Terme Femminili (?)
ID 09 TF 104	Pom-134	Cranio	Terme Femminili (?)
ID 10 17, TF33	Pom-135	Cranio	Terme Femminili (?)
ID 11 62, TF32	Pom-136	Cranio	Terme Femminili (?)
ID 13 TF42	Pom-137	Cranio	Terme Femminili (?)
ID 14 TF117	Pom-138	Cranio	Terme Femminili (?)
ID 15 TF68	Pom-139	Cranio	Terme Femminili (?)
ID 16 26, TF118	Pom-140	Cranio	Terme Femminili (?)
ID 17 115, TF54	Pom-141	Cranio	Terme Femminili (?)
ID 18	Pom-142	Cranio	Terme Femminili (?)
ID 19 51, TF 57	Pom-143	Cranio	Terme Femminili (?)
ID 20 6, TF16	Pom-144	Cranio	Terme Femminili (?)
ID 22 3, TF50	Pom-145	Cranio	Terme Femminili (?)

ID Archeologico	ID Laboratorio	Elemento scheletrico	Sito Archeologico
ID 23 48, TF65	Pom-146	Cranio	Terme Femminili (?)
ID 24 33, TF69	Pom-147	Cranio	Terme Femminili (?)
ID 26 110, TF4	Pom-148	Cranio	Terme Femminili (?)
ID 28 TF77, S5	Pom-149	Cranio	Terme Femminili (?)
ID 29 86, TF94	Pom-150	Cranio	Terme Femminili (?)
ID 30 17 TF78	Pom-151	Cranio	Terme Femminili (?)
ID 32 TF27	Pom-152	Cranio	Terme Femminili (?)
ID 33 39, TF76	Pom-153	Cranio	Terme Femminili (?)
ID 34 TF63	Pom-154	Cranio	Terme Femminili (?)
ID 35 39, TF109	Pom-155	Cranio	Terme Femminili (?)
ID 36 TF30	Pom-156	Cranio	Terme Femminili (?)
ID 37 TF99	Pom-157	Cranio	Terme Femminili (?)
Subadulto 1	Pom-158	Rocca petrosa sx	Casa del primo piano
Subadulto 2	Pom-159	Rocca petrosa sx	Casa del primo piano
Subadulto 3	Pom-160	Rocca petrosa sx	Casa del primo piano
Bronze child	Pom-161	Rocca petrosa sx	Casa del primo piano
Adulto 1	Pom-162	Rocca petrosa sx	Casa del primo piano
Denti rosa (femmina 1)	Pom-163	1° molare inf. dx	Casa del primo piano
Femmina 2	Pom-164	2° molare inf. dx	Casa del primo piano
Uomo adulto	Pom-165A	Epistrofeo	Casa del primo piano
Uomo adulto	Pom-165B	Prima falange	Casa del primo piano
Regio I,8,14	Pom-166	Rocca petrosa	Regio I, Cavaliere cavallo maiuri

Tabella 1 Resti umani recuperati dal Parco Archeologico di Pompei e processati per le analisi genomiche.

Tutti gli individui in *Tabella 1* sono stati processati e sequenziati fatta eccezione per Pom-8.1, Pom-8.2, Pom-10.1 (III1 II7), Pom-95B e Pom-97 che sono in fase di campionamento, mentre Pom-138, Pom-141, Pom-142, Pom-148, Pom-149, Pom-150, Pom-151, Pom-152, Pom-153, Pom-154, Pom-155, Pom-156, Pom-157 sono prossimi al sequenziamento.

3.2 Caso studio: La Stanza degli Scheletri

Tra i numerosi resti umani recuperati nell'antica città di Pompei, un caso particolare è rappresentato dagli individui ritrovati all'interno della cosiddetta "Stanza degli Scheletri" facente parte dell'edificio nella Regio V che prende il nome di "Casa del Giardino". Nel corso dei secoli, il sito è stato interessato da numerosi scavi, già prima del periodo Borbonico, i quali hanno visto l'intervento di tombaroli alla ricerca di materiali e oggetti preziosi che gli abitanti della città tenevano con sé nel corso dell'eruzione sperando in un salvataggio. Come già accennato precedentemente, poiché gli scavi iniziati nel 1748 non avevano come obiettivo il recupero e la valorizzazione delle opere e delle strutture urbanistiche della città, molti ambienti non hanno mantenuto l'assetto successivo alla tragedia. Infatti, a causa di questo comportamento incauto la "Stanza degli Scheletri" mostra evidenti segni di manomissione come lo sparpagliamento delle ossa e la presenza di fori nella parete dell'ambiente. Ciononostante, sono stati ritrovati resti umani appartenenti apparentemente a 12 individui diversi (vedi in *Tabella 1*

individui da Pom-1 a Pom-12) e tra di essi, ancora in situ, un cranio schiacciato da una delle tegole del tetto crollato in seguito a una delle correnti piroclastiche (Figura 8).

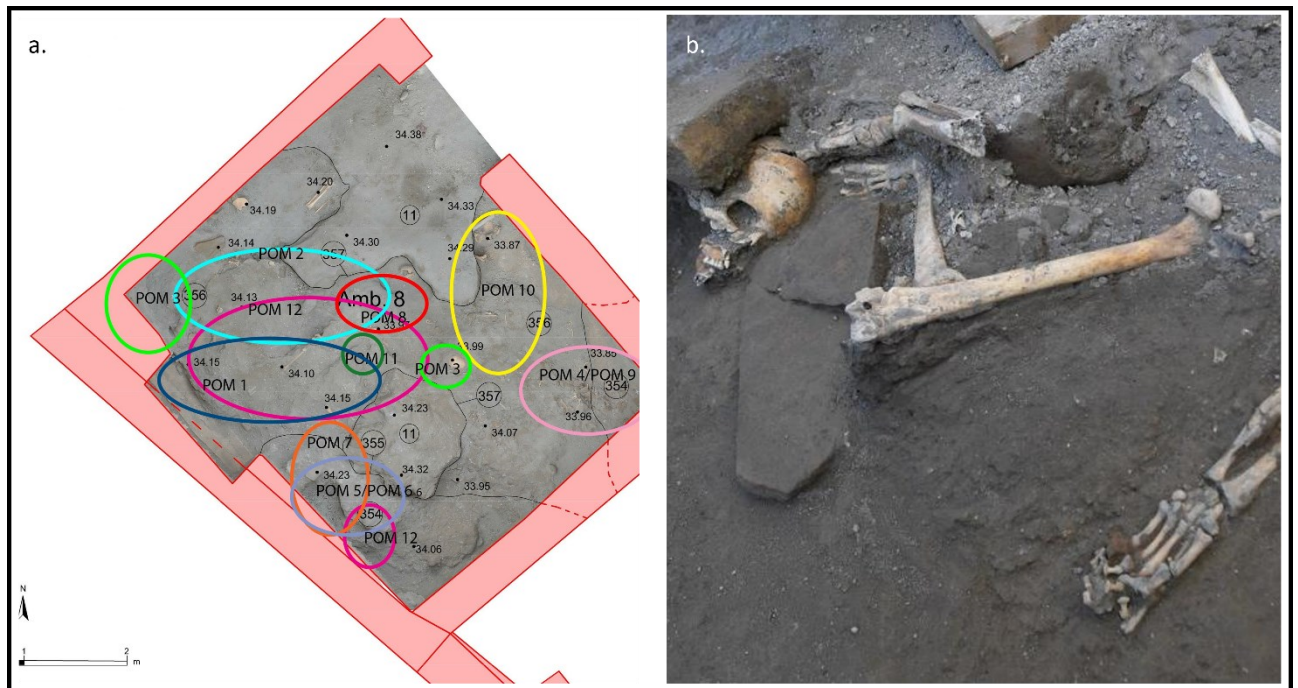


Figura 8 a. Planimetria della “Stanza degli Scheletri” in cui è evidenziata la posizione di ritrovamento dei resti umani. **b.** Cranio appartenente a uno degli adulti all’interno della stanza morto a causa del crollo del tetto (immagine estratta dal sito ufficiale del Parco Archeologico di Pompei).

La “Stanza degli Scheletri” rappresenta un caso particolare dello studio poiché all’interno di essa, nel disperato tentativo di trovarvi rifugio, persero la vita adulti e bambini probabilmente abitanti della *domus* e forse legati da parentele. Interessante, nonostante le diverse manomissioni del sito, fu il ritrovamento degli adulti in prossimità della porta (Pom-2, Pom-3, Pom-10 e Pom-12) e dei subadulti verso il fondo della stanza (Pom-1, Pom-4, Pom-5, Pom-6, Pom-7, Pom-8, Pom-9, Pom-11) come è possibile osservare in *Figura 8 a*.

3.3 Metodi di laboratorio

I resti umani oggetto dello studio sono stati processati nel laboratorio di Antropologia Molecolare e Paleogenomica dell'Università degli Studi di Firenze, in laboratorio dedicato esclusivamente al DNA antico e applicando rigorosi criteri per la prevenzione di contaminazioni in tutte le fasi degli esperimenti (Gilbert et al., 2005; Willerslev & Cooper, 2005). Inoltre, per monitorare eventuali contaminazioni durante l'estrazione del DNA e la preparazione delle librerie, sono stati inclusi dei controlli negativi.

Nel laboratorio in cui viene manipolato il DNA antico gli ambienti di lavoro sono suddivisi in una parte pre-amplificazione del DNA dedicata alla pulizia, al campionamento, all'estrazione e alla preparazione delle librerie genomiche e in una parte post-amplificazione in cui le molecole di aDNA, già sottoposte a PCR (Reazione a Catena della Polimerasi), non sono più suscettibili alle contaminazioni esterne. L'accesso al laboratorio è permesso ai soli operatori provvisti di DPI (Dispositivi di Protezione Individuale) per impedire che avvenga contaminazione con DNA moderno.

Il campionamento di tutti i resti umani, necessario a ottenere materiale sufficiente per l'estrazione dell'aDNA, è avvenuto previa documentazione fotografica e autorizzazione da parte del Parco Archeologico di Pompei.

Il materiale umano di partenza è costituito da rocche petrose (disarticolate e non) nelle quali il DNA endogeno risulta meglio preservato (Gamba et al., 2014; Pinhasi et al., 2015), da denti e da resti post-craniali (*Figura 9*). La pulizia degli elementi ossei ha come obiettivo la rimozione dei residui terrosi e precede una fase di irradiazione con luce ultravioletta (254 nm) per 45 minuti all'interno del Biolink DNA Crosslinker (BiometraTM) per degradare contaminanti di origine ambientale. In alcuni casi (Pom-56, Pom-62, Pom-67, Pom-105, Pom-106, Pom-107, Pom-118, Pom-124, Pom-127, Pom-128, Pom-138, Pom-142, Pom-151, Pom-158, Pom-160, Pom-162, Pom-166), la pulizia del meato acustico esterno situato nella rocca ha consentito il recupero di tutti o parte degli ossicini dell'udito quali staffa, incudine e martello. Quest'ultimi sono stati utilizzati nella loro interezza per l'estrazione del DNA (Sirak et al., 2020) successivamente a una pulizia superficiale con acqua distillata ed esposizione alla radiazione ultravioletta per 10 minuti.

Le rocche petrose disarticolate sono state tagliate longitudinalmente con lame diamantate circolari montate su un trapano dentistico per consentire l'esposizione della zona densa del tessuto osseo (*Figura 10*). La polverizzazione è stata svolta tramite l'uso di microfresse diamantate e rotanti, su trapano dentistico, a bassa velocità per non surriscaldare la zona di campionamento. Da ciascuna rocca sono stati ricavati circa 50 mg di polvere in prossimità della zona c. In caso di rocca petrosa articolata nel cranio (Pom-96B, Pom-99, Pom-100, Pom-101, Pom-102, Pom-109, Pom-110, Pom-112, Pom-116, Pom-117, Pom-119, Pom-122, Pom-125, Pom-129, Pom-130, Pom-131, Pom-132, Pom-133, Pom-134, Pom-135, Pom-137, Pom-139, Pom-140, Pom-141, Pom-143, Pom-144, Pom-145, Pom-146, Pom-147, Pom-148, Pom-149, Pom-150, Pom-152, Pom-153, Pom-154, Pom-155, Pom-156, Pom-157), per

garantire la preservazione del campione, è stata utilizzata una strategia meno invasiva di microcampionamento (Sirak et al., 2017). Essa prevede la polverizzazione della zona c, tramite l'impiego di microfresse diamantate, accedendo dalla fossa giugulare o a livello della cresta che divide quest'ultima dal canale carotideo.

Dai denti i 50 mg di polvere sono stati ricavati campionando all'interno della radice.



Figura 9 Alcuni dei resti umani utilizzati per l'analisi paleogenomica. In alto: rocche petrose articolate nel cranio e rocca sx isolata. In basso, da sinistra verso destra, incudine, primo premolare inferiore sx e falange.

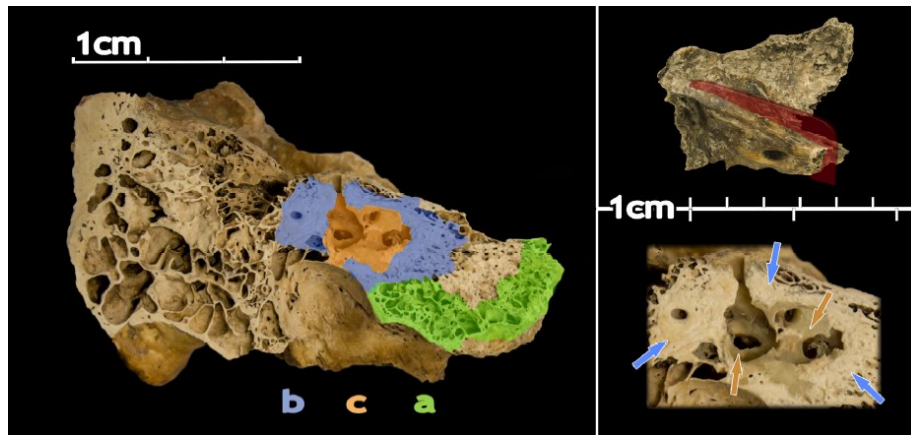


Figura 10 Piano mediale di una rocca sx sezionata. Nell'immagine principale sono evidenziate le aree analizzate; la zona da cui prelevare la polvere per l'estrazione del DNA è segnata in arancione (zona c). L'immagine in alto mostra il piano di taglio della rocca, mentre in basso è possibile osservare il dettaglio delle zone c e b. (immagine estratta da Pinhasi et al., 2015).

Il protocollo più comune utilizzato per l'estrazione del DNA dalla polvere, ottenuta a partire dai resti ossei e dai denti, è quello messo a punto da Rohland e Hofreiter nel 2007. Nel 2013 il protocollo venne ottimizzato da Dabney et al. (2013) con l'obiettivo di recuperare corti frammenti di acidi nucleici quali quelli di aDNA. L'estrazione del DNA viene realizzata in due giornate e prevede inizialmente l'utilizzo di un *extraction buffer* composto da EDTA (0.45M, pH 8) e proteinasi K (0.25 mg/mL). La polvere di ciascun campione viene sospesa in 1 mL di *extraction buffer* e incubata per l'intera notte a 37°C su un supporto rotante. Il secondo giorno, dopo aver centrifugato la sospensione, il surnatante viene recuperato e mescolato al *binding buffer* composto da cloridrato di guanidina, sodio acetato e isopropanolo. Esso consente ai frammenti di DNA di legarsi alla membrana in silica contenuta all'interno di apposite falcon (High Pure Viral Acid Large Volume Kit, Roche) riempite della miscela. Dopo aver eseguito due cicli di lavaggio con un *wash buffer* a base di etanolo, il DNA viene eluito in due ripetizioni successive utilizzando 50 µL di *elution buffer* (10 mM Tris-Cl, pH 8.5) ottenendo un totale di 100 µL di estratto per campione.

La concentrazione degli estratti viene valutata tramite Qubit 3.0 Fluorometer utilizzando il kit Qubit dsDNA High Sensitivity (Thermo Fischer Scientific).

30 µL di estratto sono utilizzati per la costruzione delle librerie Illumina a doppio filamento seguendo un protocollo studiato appositamente per il DNA antico (Meyer and Kircher, 2010). Durante la preparazione delle librerie i frammenti di DNA vengono sottoposti a trattamento *half-UDG* (Rohland et al., 2015) tramite l'impiego del complesso enzimatico USER (*Uracil-Specific Excision Reagent*); esso è in grado di riconoscere e rimuovere le citosine deamminate, abbondanti nelle molecole di aDNA. La sua azione viene successivamente inibita tramite l'impiego dell'enzima UGI (*Uracil Glycosylase Inhibitor*), il

quale fa sì che le deamminazioni vengano limitate alle ultime due paia di basi in 5' e 3' al fine di consentire l'autenticazione del dato antico.

L'*indexing* delle librerie è effettuata preparando una mix di reazione contenente la Pfu Turbo DNA Polymerase e utilizzando, per ciascun campione, una coppia univoca di corte sequenze oligonucleotidiche (8 bp); il legame degli indici avviene impostando 15 cicli di PCR.

Le librerie indicizzate vengono poi purificate utilizzando le colonnine MinElute (Qiagen) e il DNA infine eluito in 20 µL di *elution buffer*. Tramite corsa elettroforetica su *Tapestation System Agilent 4150*, sono valutate qualitativamente e quantitativamente le librerie utilizzando una D1000 ScreenTape (per frammenti da 35 a 1000 bp). Nei casi in cui il profilo delle librerie mostra la presenza di molecole eteroduplici, strutture ibride di DNA costituite da filamenti non complementari appaiati, si rende necessario effettuare 2 cicli di PCR preparando una mix di reazione contenente l'Herculase II Fusion DNA Polymerase (*Agilent Technologies*) e un set di primers IS5-IS6. Terminata la reazione, le librerie vengono nuovamente purificate tramite l'impiego delle colonnine MinElute e il DNA eluito il 20 µL di *elution buffer*. La concentrazione del DNA e la verifica dell'effettiva rimozione degli eteroduplici viene valutata sempre utilizzando la *Tapestation System Agilent 4150* con kit D1000.

Per questo lavoro tutte le librerie sono state diluite a 4 nM e usate per la costruzione di un unico *pool* di sequenziamento per la fase di *screening*, con l'obiettivo di verificare la presenza del DNA endogeno. Per questo primo sequenziamento, di tipo *shotgun light*, sono state generate 5 milioni di *reads* per campione. Il sequenziamento su piattaforma Illumina NovaSeq 6000 (kit 2x50 e 2x100 cicli) è stato effettuato presso il Laboratorio di Genomica Avanzata del Dipartimento di Biologia dell'Università degli Studi di Firenze.

3.4 Analisi bioinformatica

3.4.1 Trattamento dei dati grezzi – *shotgun* esplorativo

Effettuato il *demultiplexing*, i primi dati di sequenziamento in formato fastq sono stati analizzati tramite l'impiego del *software* EAGER v1.92.59 (Peltzer et al., 2016) al fine di valutare la preservazione del materiale genetico endogeno. Questo software è stato appositamente studiato per la ricostruzione di genomi antichi ed è in grado, partendo dalle *reads* grezze, di produrre un allineamento in formato BAM (*Binary Alignment Map*). All'interno del *software* sono contenuti differenti strumenti per il *pre-processing* delle *reads*: tramite FastQC viene controllata la qualità dei frammenti di DNA; la rimozione delle sequenze degli adattatori Illumina e l'unione delle *reads forward* e *reverse* viene invece effettuata da Clip&Merge settando la minima sovrapposizione delle reads a 10 bp in caso di sequenziamento con kit 2x100 (in modo da recuperare solo frammenti di piccole dimensioni ed evitare di trattenere molecole contaminanti) e a 1 bp in caso di kit 2x50. Infine, le *reads* più corte di 30 bp sono state scartate per impedire che venissero allineate in modo impreciso.

Le reads così filtrate sono state allineate sul genoma umano di riferimento hg19 utilizzando l'algoritmo *bwa aln* (BWA v0.7.12) (Li and Durbin, 2009) disattivando il *seeding*, impostando la qualità del mappaggio a 30 (-q 30, il che significa tollerare 1 errore ogni 1000 nucleotidi e quindi avere il 99.9% di confidenza) e consentendo il 2% di *mismatches* tra *reads* e sequenza di riferimento (-n 0.01).

Al fine di mantenere solo letture indipendenti le une dalle altre, il *software* DeDup (v0.12.2) (Peltzer et al., 2016) è stato utilizzato per rimuovere i duplicati PCR ovvero quelle reads che mappano esattamente nella stessa posizione di inizio e fine.

Il *software* mapDamage v2.0 (Jónsson et al., 2013), sempre implementato in EAGER, è stato utilizzato per verificare i *patterns* di degradazione delle molecole di aDNA e cioè piccole dimensioni dei frammenti di materiale genetico e alto tasso di deamminazione alle estremità delle molecole.

Per impedire il mantenimento delle deamminazioni alle estremità delle *reads* (non riparate dal trattamento *half-UDG* delle librerie) e di conseguenza la chiamata di falsi polimorfismi, gli allineamenti in formato BAM sono stati processati tramite il *software* BamUtil (Jun et al., 2015) usando l'opzione TrimBam. Quest'ultimo ha permesso di mascherare 2 bp, sia in 5' che in 3' delle *reads*, in modo che non fossero disponibili per i *tools* a valle.

Gli individui la cui percentuale di DNA endogeno superava il 20% e le cui molecole presentavano, sia all'estremità 5' che 3', le sostituzioni C->T e G->A caratteristiche del DNA antico sono stati selezionati per un secondo sequenziamento di tipo *shotgun* profondo (su piattaforma NovaSeq con kit 2x50 e 2x100) al fine di raggiungere una copertura media del genoma di 1x.

3.4.2 Trattamento dei dati grezzi – *shotgun* profondo

La qualità delle *reads* in formato fastq dei 12 individui selezionati per il sequenziamento *deep shotgun* e per i 6 individui della "Stanza degli Scheletri" è stata valutata utilizzando il *software* FastQC v0.12.1. L'unione delle letture *forward* e *reverse* (R1 e R2 rispettivamente) è stata poi eseguita dal *software* AdapterRemoval v2.3.3 (Schubert et al., 2016) impostando la lunghezza minima delle *reads* (--*minlength* 30) e la qualità minima delle basi a 30 (--*minquality* 30), mentre la minima sovrapposizione tra le *reads* R1 e R2 nella fase di *merging* è stata impostata a 10 in caso di sequenziamento con kit 2x100 e a 1 in caso di kit 2x50 cicli. Le *reads* sono state allineate al genoma umano di riferimento (hg19) tramite l'allineatore Burrows-Wheeler utilizzando il comando *bwa aln* v0.7.17 (Li and Durbin, 2009) specifico per il mappaggio di corte sequenze nucleotidiche (<200 bp). Per consentire l'allineamento di tutte le *reads* è stato disattivato il *seeding* (-l 16500) ed è stato inoltre concesso un *mismatch* tra letture e genoma di riferimento del 2% (-n 0.01). Le *reads* mappanti sono state filtrate per qualità di mappaggio utilizzando il *software* *samtools view* v1.16.1 (Danecek et al., 2021) il quale ha permesso di trattenere solo le letture con qualità maggiore o uguale a 30 (-q 30). Per rimuovere le *reads* mappanti sulle stesse coordinate di inizio e fine è stato impiegato il *software* DeDup v0.12.8 (Peltzer et al., 2016). La

valutazione dei *pattern* di degrado delle molecole è stata effettuata tramite il software mapDamage v2.2.1 (Jónsson et al., 2013) utilizzato anche per ricalibrare la qualità delle basi danneggiate attivando l'opzione --rescale. Il BAM file ricalibrato è stato quello utilizzato per le successive analisi.

3.4.3 Determinazione del sesso

Il sesso, su base molecolare, è stato determinato utilizzando lo script di Skoglund et al. (2013) utilizzando come input i file BAM. Esso valuta il rapporto (R_y) tra il numero di *reads* mappanti sul cromosoma Y e il numero di *reads* mappanti su entrambi i cromosomi sessuali scartando quelle con una qualità del mappaggio inferiore a 30 ($-q\ 30$). Inoltre, a ogni valore di R_y viene associato un intervallo di confidenza al 95% (CI). Se il limite inferiore dell'intervallo supera il valore di 0.077 viene assegnato all'individuo il sesso maschile, se invece il limite superiore risulta più piccolo di 0.016 viene assegnato all'individuo il sesso femminile.

3.4.4 Stima delle contaminazioni

La contaminazione negli individui di sesso maschile è stata stimata a livello nucleare sfruttando l'eterozigosi del cromosoma X (Rasmussen et al., 2011). Il possesso di un solo cromosoma X da parte dei maschi fa sì che essi risultino emizigoti. La presenza di un alto tasso di eterozigoti è quindi sinonimo di errore di sequenziamento o molto spesso di contaminazione. Per poterla quantificare è stato utilizzato il programma ANGSD v0.932 (Korneliusson et al., 2014) ed è stata presa in considerazione la stima fornita dal "*method1 new version*", effettuata sul "*Methods of Moments*" (MoM). Inoltre, sia gli individui di sesso maschile che femminile sono stati sottoposti all'analisi di stima della contaminazione a livello mitocondriale tramite lo strumento schmutzi (Renaud et al., 2015). Esso ricostruisce la sequenza consenso endogena del DNA mitocondriale considerando i *pattern* di deaminazione del campione e la distribuzione della dimensione dei frammenti e stima i livelli di contaminazione tramite il confronto con un database di frequenze alleliche di riferimento rappresentativo dell'attuale variabilità euroasiatica dell'mtDNA. Il comando ContDeam.pl è stato lanciato usando le opzioni --library double e --lengthDeam 2, in accordo con la tipologia di librerie genomiche preparate, mentre per il comando schmutzi.pl sono stati utilizzati i parametri --notusepredC e --uselength. La contaminazione è stata stimata solo per gli individui i cui dati sono stati prodotti successivamente allo *shotgun* profondo.

3.4.5 Identificazione degli aplogruppi uniparentali

Per poter identificare gli aplogruppi dell'mtDNA è stata prima ricostruita la sequenza consenso dell'individuo in esame utilizzando il comando log2fasta di schmutzi a partire dai file bam (Renaud et al., 2015) e impostando il filtro qualità a 30 ($-q30$). La sequenza in formato fasta è stata processata mediante il *tool online* Haplogrep2 v.2.4.0 (Weissensteiner et al., 2016) per l'assegnazione dell'aplogruppo.

Per quanto riguarda invece l'identificazione dell'aplogruppo del cromosoma Y (chrY) è stato utilizzato il software Yleaf v2.2 (Ralf et al., 2018) che sfrutta una serie di marcatori localizzati sul chrY e riportati all'interno del database ISOGG v.15.73 (disponibile all'indirizzo <https://isogg.org/tree/index.html>). Per l'analisi sono stati usati solamente i marcatori coperti da almeno 2 *reads* (-r 2) con qualità minima di mappaggio pari a 30 (-q 30) e considerando solamente gli alleli supportati da almeno il 90% delle *reads* (-b 90). Anche in questo caso l'analisi di identificazione degli aplogruppi è stata condotta usando i dati *deep shotgun*.

3.4.6 Genotyping e Analisi delle Componenti Principali

I dati di *shotgun* esplorativo sono stati utilizzati per analisi di tipo qualitativo e ogni individuo è stato genotipizzato mediante l'utilizzo del software *pileupCaller* (<https://github.com/stschiff/sequenceTools>). Esso è in grado di effettuare una chiamata pseudo-aploide a livello degli SNPs che caratterizzano l'*array* 1240K (1233013 polimorfismi) e *Human Origins* (HO, 597573 polimorfismi risultanti dall'unione dei *dataset* 1240K e HO). La chiamata pseudo-aploide consiste nel campionamento *random*, per ciascuna posizione, di una *read* in cui l'allele letto su di essa sarà l'approssimazione del genotipo dell'individuo. Per la chiamata, la qualità delle basi e delle *reads* mappanti è stata impostata a 30 (-q 30 e -Q 30) inoltre, è stato settato il parametro -B allo scopo di disabilitare la ricalibrazione delle basi, consigliato in caso di trattamento di dati a basso *coverage*. I dati pseudo-aploidi vengono salvati in formato *eigenstrat* caratterizzato da tre files contenenti le informazioni relative ai genotipi (.geno), agli SNPs dello *SNPs-array* utilizzato per la chiamata (.snp) e agli individui genotipizzati (.ind). I file sono stati successivamente uniti ai rispettivi dataset pubblici HO e 1240K v44.3, tramite il comando *mergeit* (EIGENSOFT v7.2.1) (Patterson et al., 2006), ai quali sono stati aggiunti altri individui antichi di confronto di recente pubblicazione (Saupe et al., 2021; Posth et al., 2021; Aneli et al., 2022; Antonio et al., 2023 bioRxiv).

È stata eseguita un'analisi delle componenti principali (PCA) tramite il programma *smartpca* (EIGENSOFT v7.2.1) utilizzando dati genomici, in formato *eigenstrat*, generati sullo *SNPs-array* HO di 61 popolazioni rappresentative della variabilità genetica dell'Eurasia occidentale e dell'Africa del nord (con l'aggiunta dei moderni toscani [TSI.SG], usati solo per il plot a causa della bassa copertura). *Smartpca* fornisce in *output* le componenti principali (*eigenvectors*) e gli *eigenvalues*. L'analisi è stata effettuata impostando il parametro *lsqproject*: YES (suggerito in caso di alta percentuale di dati mancanti) e *shrinkmode*: YES (per correggere lo stiramento degli assi causato dagli individui moderni utilizzati per la costruzione delle componenti principali che potrebbe provocare un'errata proiezione degli individui antichi, i quali tenderebbero a posizionarsi verso il centro della PCA). Sulle popolazioni moderne sono stati in seguito proiettati gli individui antichi di confronto e gli individui ritrovati a Pompei oggetto di

questo studio, di quest'ultimi sono stati però utilizzati solamente i campioni con almeno 10000 SNPs coperti sullo *SNPs-array* HO. La PCA è stata riportata in grafico utilizzando il comando ggplot di R.

3.4.7 Analisi delle parentele

La ricerca di eventuali relazioni di parentela, tra gli individui ritrovati all'interno del medesimo contesto di scavo quale la "Stanza degli Scheletri", è stata condotta utilizzando il software READ (Monroy et al., 2018). Esso è stato sviluppato per inferire relazioni di parentela, fino al secondo grado, tra coppie d'individui sequenziati a basso *coverage* (almeno 0.1x). READ utilizza come *input* dati autosomici in formato testuale e trasposto PLINK v1.9 (.tped, .tfam) (Purcell et al., 2007) ottenuti a partire dai dati di *genotyping* sullo *SNPs-array* 1240K. Per identificare le parentele READ suddivide il genoma in finestre non sovrapposte da 1 milione di bp e, per ciascuna di esse e per ciascuna coppia d'individui, calcola la proporzione di alleli differenti (P0). Il valore di P0 normalizzato per un P0 medio per coppie di individui non imparentate viene utilizzato per la classificazione in: non imparentati ($P0 \geq 0.91$), parentela di I° ($P0 \geq 0.81$), parentela di II° ($P0 \geq 0.63$) e gemelli ($P0 < 0.63$).

3.4.8 Approccio *Whole Genome* e *Genotype Likelihood*

Sui dati degli individui selezionati per il sequenziamento profondo è stato applicato un approccio di analisi di tipo *whole-genome*, il quale considera il genoma umano nella sua interezza e non per SNPs tipizzati dai pannelli precedentemente descritti.

Partendo dai bam file e tramite l'impiego di samtools depth v1.11 (Danecek et al., 2021) è stata valutata, per ciascun individuo, la copertura del genoma segnando il numero di posizioni coperte almeno 1, 3 e 5 volte. Tramite l'opzione multiinter di bedtools v2.3 (Quinlan & Hall, 2010) sono state trattenute le regioni comuni a tutti gli individui in esame e coperte anch'esse almeno 1, 3 e 5 volte. Quest'operazione è stata successivamente ripetuta per trattenere le regioni condivise anche tra percentuali sempre minori di individui.

Le analisi popolazionistiche sono state svolte considerando le regioni neutrali del genoma e cioè quelle porzioni di DNA non sottoposte all'azione della selezione naturale. Per trattenere le regioni neutrali sono state applicate sul genoma di ciascun individuo delle maschere utilizzando l'opzione subtract di bedtools v2.3 (Quinlan & Hall, 2010). In questo modo sono state escluse le sequenze di DNA codificante (CDS), le sequenze 10000 bp a monte e a valle dei geni, le regioni ripetute (elementi trasponibili LTR e non-LTR come le sequenze LINE e SINE, sequenze satelliti ecc) e le isole CpG.

Con un *coverage* medio del genoma di 1x la strategia adottata per identificare le varianti genomiche non è stata la chiamata diretta del genotipo, bensì il calcolo delle *genotype likelihoods* (GL).

L'esigenza di sviluppare modelli probabilistici per usare le GL negli studi di genetica di popolazione nasce dal notevole incremento nella produzione di dati NGS (*Next Generation Sequencing*) a basso

coverage. Essi rappresentano però una sfida in fase di chiamata dei genotipi poiché, a causa delle scarse informazioni, non si è in grado di discernere gli errori di sequenziamento dalle reali varianti genomiche. La *genotype likelihood* è un indice di verosimiglianza, indica quanto verosimile è ciascuno dei 10 possibili genotipi (AA, AT, AC, AG, CC, CT, CG, GG, GT, TT) in funzione dei dati che sono stati osservati. Di conseguenza a ogni genotipo viene associata un'incertezza che dipende dal numero di letture dell'allele o degli alleli, dalla qualità di lettura delle basi e dalla qualità del mappaggio. In caso di *missing-data* ogni genotipo ha perciò la stessa probabilità di essere osservato che equivale a 1/10. Per il calcolo delle GL è stato utilizzato il *software* ANGSD v0.932 (Korneliussen et al., 2014), sviluppato per analizzare dati NGS. È stato scelto il modello GATK (McKenna et al., 2010) (-GL 2) per il calcolo delle GL e da quest'ultime sono stati identificati gli alleli *major* e *minor* (-doMajorMinor 1), le frequenze alleliche sono state calcolate con l'opzione -doMaf 1 fissando i due alleli (per singolo locus l'allele letto meno frequentemente è l'allele *minor*), infine sono state scartate le basi e le *reads* mappanti con qualità inferiore a 30 (-minQ 30 -minMapQ 30). È stato inoltre settato il parametro -SNP_pval 1e-6 per definire un valore soglia oltre il quale il marcatore può essere considerato polimorfico e quindi incluso nell'analisi. Al *software*, tramite l'opzione -bam, è stato fornito un file testuale contenente i *path* ai bam file degli individui oggetto dello studio e di tutti i campioni antichi di confronto. Le GL sono state salvate in formato *beagle* (-doGlf 2), richiesto per le analisi a valle. Le likelihood sono state calcolate sui siti neutrali delle regioni autosomiche del genoma (escludendo quindi i cromosomi sessuali e il genoma mitocondriale) e sulle posizioni, coperte almeno una volta, comuni all'80% degli individui (focus dello studio) fornendo al *software* un file testuale con l'opzione -sites.

3.4.8.1 Selezione degli individui di confronto

Per poter studiare la variabilità genomica degli individui ritrovati a Pompei e poterla confrontare con quella di altre popolazioni, sia del territorio dell'Impero Romano che delle sue province, analizzandone anche i cambiamenti nel corso del tempo è stato costruito un dataset di popolazioni antiche già studiate e quindi presenti in letteratura. Inoltre, al fine di identificare nel genoma degli individui in esame i principali contributi genetici ancestrali che meglio descrivono la loro variabilità, sono stati selezionati dalla letteratura 10 genomi attribuibili a 5 popolazioni antiche: cacciatori-raccoglitori dell'ovest e del Caucaso (WHG e CHG dall'inglese *Western Hunter-Gatherer* e *Caucasian Hunter-Gatherer*), primi agricoltori Neolitici dell'Anatolia, agricoltori Neolitici Iraniani e pastori nomadi delle Steppe. Nel dataset sono stati inclusi gli individui il cui genoma è stato letto con un approccio di tipo *shotgun* raggiungendo un *coverage* medio maggiore o uguale a 0.1x. Le popolazioni selezionate coprono un transetto temporale che va quindi dal Mesolitico fino al periodo Imperiale, quest'ultimo comprendente non solo il territorio Italiano, ma anche quello Inglese, Francese e Portoghese sedi di colonie dell'Impero. Le popolazioni incluse all'interno del dataset con i relativi riferimenti sono riportati in *Tabella 2*. I genomi

degli individui usati per il confronto sono stati scaricati dagli archivi genomici e processati utilizzando la *pipeline* bioinformatica usata per i dati *deep shotgun* di Pompei al fine di uniformare i dati ed evitare bias causati dall'utilizzo di genomi di riferimento diversi o da un diverso trattamento dei dati grezzi.

ID popolazione	Pubblicazione
WHG	Lazaridis et al., 2014
WHG	Jones et al., 2015
CHG	Jones et al., 2015
CHG	Jones et al., 2015; Maisano Delser et al., 2021
Iran_N	Broushaki et al., 2016
Anatolia_N	Hofmanova et al., 2016
Anatolia_N	Kilinc et al., 2016
Steppe	de Barros Damgaard et al., 2018
Steppe	Allentoft et al., 2015
Italy_CA	Antonio et al., 2019
Italy_CA	Saupe et al., 2021
Italy_BA	Saupe et al., 2021
Italy_IA	Antonio et al., 2019
Italy_IA	Posth et al., 2021
Italy_IA	Aneli et al., 2022
Portugal_Imperial	Antonio et al., 2023 bioRxiv
France_Imperial	Antonio et al., 2023 bioRxiv
England_Roman	Martiniano et al., 2016
Italy_Imperial	Antonio et al., 2019
Italy_Imperial	Posth et al., 2021
Italy_Imperial	Antonio et al., 2023 bioRxiv

Tabella 2 ID delle popolazioni usate nelle analisi *whole-genome* con relativi riferimenti bibliografici.

3.4.8.2 PCAngsd

L'analisi delle componenti principali (PCA) sui genomi con *coverage* medio 1x è stata condotta utilizzando il programma PCAngsd, adeguato a dati NGS a medio e basso *coverage* sfruttando le GL (Meisner & Albrechtsen, 2018). PCAngsd usa come input le GL in formato *beagle* e fornisce come output una matrice di covarianza. Gli *eigenvalues* e gli *eigenvectors* della matrice di covarianza sono stati calcolati utilizzando la funzione `eigen()` di R tramite il quale, con il comando `plot()`, sono portati in grafico i valori degli *eigenvectors*.

3.4.8.3 NgsAdmix

La proporzione individuale di *admixture*, sempre per gli individui *deep shotgun*, è stata stimata utilizzando lo strumento NgsAdmix. Anch'esso sfrutta le GL calcolate a partire da dati NGS a basso *coverage* (Skotte et al., 2013; Meisner & Albrechtsen, 2018). NgsAdmix è in grado di rilevare segnali recenti di *admixture* che hanno generato variazione nelle frequenze alleliche della popolazione. Il

programma utilizza in *input* GL in formato *beagle* e sono stati generati diversi file di *output* variando il numero di popolazioni sorgente da 2 a 10 (-K 2 ÷ -K 10). Infine, è stata settata al 5% la minima frequenza allelica (-minMaf 0.05 parametro di default) sotto la quale le varianti, considerate rare, sono state scartate dall'analisi. Il file di *output* con opportuno K e con estensione qopt è stato utilizzato per generare il grafico finale creato tramite il pacchetto pophelper di R.

4 RISULTATI E DISCUSSIONI

4.1 Shotgun esplorativo

Il sequenziamento a 5 milioni di *reads* delle 150 librerie ha prodotto delle sequenze grezze la cui analisi, tramite il *software* Eager v.1.92.59 (Peltzer et al., 2016), ha restituito le statistiche principali riportate in *Tabella 3*.

I risultati ottenuti tramite l'impiego del *software* mapDamage v2.0 (Jónsson et al., 2013) al fine di autenticare le molecole di DNA antico tramite i loro pattern di degradazione sono riportati in *Tabella 4* e in *Figura 11*.

Nella *Tabella 3* sono presenti, per ciascun individuo, il numero di *reads* grezze generate, il numero di *reads* mappati dopo la rimozione dei duplicati PCR e l'applicazione dei filtri qualità, la percentuale di DNA endogeno e il *cluster factor* (rapporto tra la quantità di *reads* prima e dopo la rimozione dei duplicati PCR).

ID campioni	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Cluster factor
Pom-1	17334740	1497783	27.768	0.0252	1.165
Pom-2	11768430	480247	13.072	0.0079	1.131
Pom-3	9216562	1468742	46.02	0.0254	1.002
Pom-4 /Pom-9	8914330	1829616	62.912	0.0295	1.002
Pom-5 /Pom-6	7909734	1487789	57.732	0.0237	1.002
Pom-7	7708392	148939	7.015	0.0018	1.002
Pom-8	17540314	11359	0.252	0.0002	1.154
Pom-10	15986928	23326	0.526	0.0004	1.153
Pom-10.1	18692022	7073	5.356	0.0001	1.296
Pom-11	8733370	278	0.019	0	1
Pom-11.1	12131420	7563	1.754	0.0001	1.116
Pom-12	14105318	1486143	35.452	0.0259	1.145
Pom-13	6994790	2670	1.051	0	1.245
Pom-15	9989270	1913	2.435	0	1.064
Pom-16A	16980382	293664	9.125	0.0053	1.319
Pom-16B	15842786	238144	9.099	0.005	1.646
Pom-17	4596580	272	0.045	0	1.011
Pom-18	17632346	783	0.212	0	1.203
Pom-19	5381742	208716	9.293	0.0041	1.004
Pom-20	16939576	3166144	56.973	0.0526	1.2
Pom-21	31584038	1295350	12.746	0.0198	1.003
Pom-22	23499842	8587	0.473	0.0001	1.221
Pom-23	14557696	11546	4.489	0.0002	1.143
Pom-24	14664654	1071115	25.747	0.0156	1.104
Pom-25	12513118	968	2.674	0	1.257
Pom-26	11590940	2138653	59.253	0.0331	1.090
Pom-27	12035142	281362	8.95	0.0041	1.093
Pom-28	42887336	1280897	12.528	0.0191	1.098
Pom-29	14191600	6667	1.352	0.0001	1.092
Pom-30	12305416	2694582	73.81	0.0406	1.096

ID campioni	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Cluster factor
Pom-31	13830112	71408	2.472	0.001	1.110
Pom-32	10700412	1075558	36.911	0.0152	1.088
Pom-33	11465940	72633	2.163	0.0011	1.082
Pom-34	10424204	712	0.572	0	1.103
Pom-35	12732576	93449	2.345	0.0015	1.097
Pom-36	11409410	2565355	64.927	0.0405	1.099
Pom-37	8623328	1164378	36.49	0.0175	1.085
Pom-38	11446984	12605	0.653	0.0002	1.093
Pom-39	10710536	34014	1.317	0.0005	1.085
Pom-40	12665066	1890823	44.558	0.0292	1.095
Pom-41	14669830	2915042	57.652	0.0458	1.082
Pom-43	13170980	27152	1.446	0.0004	1.089
Pom-44	11199616	37804	1.413	0.0005	1.091
Pom-45	14185672	314918	7.596	0.0044	1.102
Pom-46	13888694	57695	3.234	0.0008	1.083
Pom-47	39712992	59386	0.649	0.0008	1.091
Pom-48	57686026	11009996	63.529	0.1699	1.094
Pom-49	51524626	6066	1.951	0.0001	1.286
Pom-50	50489070	843390	4.827	0.0136	1.102
Pom-51	46306698	183079	1.584	0.0027	1.099
Pom-52	48646526	3800	0.218	0	1.085
Pom-53	78956214	629746	2.75	0.0085	1.101
Pom-54	94439724	15322385	47.402	0.2633	1.112
Pom-55	44338156	303169	2.072	0.0046	1.096
Pom-56	57577900	7078821	45.215	0.1119	1.095
Pom-57	63361320	2117	0.255	0	1.119
Pom-58	57919990	109649	0.847	0.0015	1.09
Pom-59	31008314	2458	0.275	0	1.048
Pom-60	47488452	5848	1.102	0.0001	1.193
Pom-61	77114288	1394360	8.384	0.0203	1.093
Pom-62	53807276	17097941	88.253	0.2826	1.096
Pom-63	45524634	2291114	15.223	0.0351	1.099
Pom-64	64585610	6640361	33.014	0.1038	1.083
Pom-65	92268444	6677	2.739	0.0001	1.564
Pom-66	54190440	6108488	34.134	0.0938	1.086
Pom-67	54451340	446193	3.158	0.006	1.089
Pom-68	68164770	47371	0.606	0.0007	1.094
Pom-69	53032428	939169	4.991	0.0147	1.095
Pom-70	9415338	42084	2.329	0.0006	1.089
Pom-71	9641348	21007	1.195	0.0003	1.084
Pom-72	28323050	1349244	22.756	0.0196	1.098
Pom-73	9107130	106785	5.279	0.0018	1.085
Pom-74	16902258	6046	0.173	0.0001	1.079
Pom-75	12723178	434	0.031	0	1.101
Pom-76	12010472	113050	5.055	0.0017	1.094
Pom-77	12421892	2977461	94.274	0.05	1.091
Pom-78	47639490	3325941	28.391	0.0519	1.088
Pom-79	65033064	1720237	16.68	0.0271	1.088
Pom-80	24440610	831124	15.329	0.013	1.091

ID campioni	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Cluster factor
Pom-81	24383514	2796	0.078	0	1.099
Pom-82	11336616	30267	1.225	0.0005	1.095
Pom-83	12255478	2705	0.106	0	1.082
Pom-84	10583396	5792	0.282	0.0001	1.083
Pom-85	18360804	316614	10.854	0.0042	1.152
Pom-86	9694000	36703	1.825	0.0005	1.085
Pom-87	63915098	2041850	11.056	0.0324	1.091
Pom-88	13510586	2762788	80.422	0.0417	1.106
Pom-89	17511850	936	0.039	0	1.062
Pom-90	32985930	2485058	32.596	0.037	1.082
Pom-91	9367612	22294	1.376	0.0003	1.091
Pom-92	16280860	436	0.035	0	1.188
Pom-93	14217864	469	0.034	0	1.102
Pom-96B	25816846	2551546	30.217	0.044	1.121
Pom-98	30332192	3653	0.625	0.0001	1.063
Pom-99	32414444	610501	6.553	0.0086	1.114
Pom-100	28320056	1531729	16.466	0.023	1.103
Pom-101	19441518	831102	13.58	0.0125	1.103
Pom-102	25662206	243631	3.572	0.0037	1.11
Pom-103	18742766	2638519	40.943	0.0403	1.116
Pom-104	25714800	2823963	32.039	0.0456	1.109
Pom-105	14792338	5283816	95.14	0.0894	1.124
Pom-106	16123890	1489629	29.141	0.0228	1.103
Pom-107	16750040	5022464	93.065	0.0798	1.107
Pom-108	12925002	393006	8.114	0.0062	1.1
Pom-109	18456868	20744	0.846	0.0003	1.095
Pom-110	18823670	1785076	28.647	0.0254	1.1
Pom-111	14448664	743047	15.251	0.0116	1.096
Pom-112	15775090	1573613	29.961	0.0241	1.101
Pom-113	44710218	4076683	27.388	0.0651	1.122
Pom-114	21706264	247314	3.928	0.0038	1.112
Pom-115	19075872	60050	1.421	0.0009	1.107
Pom-116	15486616	25655	0.859	0.0004	1.119
Pom-117	19928620	1018687	17.726	0.0146	1.108
Pom-118	19023456	6053789	91.8	0.1006	1.116
Pom-119	21250038	60273	2.439	0.0009	1.111
Pom-120	45037942	200546	2.042	0.0031	1.128
Pom-121	15453040	876064	17.709	0.0127	1.111
Pom-122	20057744	1554647	24.395	0.0242	1.129
Pom-123	23160800	5183246	63.048	0.0842	1.114
Pom-124	39458138	11721614	94.769	0.1817	1.11
Pom-125	19251600	2873361	50.227	0.0498	1.147
Pom-126	15694056	2085721	39.791	0.0327	1.103
Pom-127	11874886	2544574	81.948	0.0351	1.174
Pom-128	17018716	3176361	73.981	0.0535	1.189
Pom-129	7899384	1317797	50.526	0.0213	1.039
Pom-130	8766362	33646	2.302	0.0005	1.04
Pom-131	11729364	384652	10.598	0.0059	1.051
Pom-132	7377614	1493994	61.161	0.026	1.042

ID campioni	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Cluster factor
Pom-133	6039136	97495	5.077	0.0016	1.042
Pom-134	3081374	88239	8.483	0.0014	1.035
Pom-135	8661206	133448	5.93	0.0019	1.05
Pom-136	12368640	618208	17.86	0.0093	1.041
Pom-137	11321984	316806	9.01	0.0054	1.036
Pom-139	4955446	444589	27.178	0.0076	1.038
Pom-140	11151332	596625	19.928	0.0107	1.144
Pom-143	14324414	183621	7.498	0.0028	1.137
Pom-144	10318860	381735	13.093	0.0054	1.139
Pom-145	10952110	107821	10.216	0.0016	1.149
Pom-146	9885504	182230	12.149	0.0024	1.135
Pom-147	10463740	7089	5.586	0.0001	1.208
Pom-158	18006246	1212762	22.421	0.0187	1.123
Pom-159	16808038	1564109	32.124	0.0253	1.121
Pom-160	16880072	3537450	64.756	0.0618	1.104
Pom-161	26435022	4735259	58.602	0.0782	1.121
Pom-162	15037490	3831389	86.9	0.0594	1.129
Pom-163	11349740	169233	5.79	0.0026	1.121
Pom-164	17559342	66184	1.943	0.0012	1.118
Pom-165A	12671210	102128	7.323	0.0016	1.142
Pom-165B	20043694	400015	6.826	0.0068	1.115
Pom-166	15948184	384314	10.038	0.0059	1.107

Tabella 3 Principali statistiche relative al sequenziamento esplorativo dei 150 campioni processati. In rosso sono evidenziati gli individui selezionati per il sequenziamento deep shotgun.

La percentuale di DNA endogeno (frazione di *reads* mappanti sul genoma umano di riferimento rispetto al totale di *reads* ottenute dal *merging*) è stato il parametro che ha avuto maggior peso nella scelta delle librerie da sottoporre a un secondo sequenziamento di tipo *deep shotgun*, con l'obiettivo di raggiungere un *coverage* medio di 1x. I valori di DNA endogeno variano da un minimo di 0.019% a un massimo di 95.14%. La preservazione del DNA risulta quindi molto variabile: nel 67% dei campioni si osserva una percentuale di DNA inferiore al 20%, mentre il 33% dei campioni mostra una preservazione superiore al 20%; risultati eccellenti considerata la natura degradata dell'aDNA. In tutti i campioni vi è un basso numero di duplicati PCR rispetto al totale di *reads* uniche mappanti sul genoma umano di riferimento, dimostrato dal valore del *cluster factor* che si aggira attorno all'1 indicando un'elevata complessità delle librerie.

Nella *Tabella 4* sono riportate, per ciascun campione, le percentuali di transizioni C->T e G->A del primo nucleotide rispettivamente alle estremità 5' e 3' delle molecole di DNA e la dimensione media dei frammenti.

Le percentuali di C->T variano da 0.77% (in Pom-93 a causa della bassa percentuale di DNA endogeno) a 27.67%, mentre i valori per le transizioni G->A variano da 2.4% a 26.1%. L'incidenza delle misincorporazioni, sia in 5' che in 3', è compatibile con l'antichità dei resti umani e con il trattamento

half-UDG applicato in fase di costruzione delle librerie. Anche la dimensione media dei frammenti, tra 30 a 62 paia di basi, è coerente con quanto generalmente osservato nei campioni di aDNA.

ID campione	Deamminazione in 5' (%)	Deamminazione in 3' (%)	Dimensione media dei frammenti
Pom-1	0.1114	0.1084	47
Pom-2	0.1066	0.1048	45
Pom-3	0.0727	0.0714	48
Pom-4 /Pom-9	0.0642	0.0637	45
Pom-5 /Pom-6	0.0631	0.0612	44
Pom-7	0.1222	0.1261	37
Pom-8	0.0959	0.0906	46
Pom-10	0.0495	0.0471	51
Pom-10.1	0.0891	0.0652	41
Pom-11	0.0194	0.0476	31
Pom-11.1	0.068	0.065	43
Pom-12	0.1165	0.1065	48
Pom-13	0.0608	0.0633	47
Pom-15	0.149	0.1017	33
Pom-16A	0.1429	0.1397	52
Pom-16B	0.051	0.0495	62
Pom-17	0.038	0.0116	31
Pom-18	0.0283	0.0218	30
Pom-19	0.0439	0.0428	56
Pom-20	0.0927	0.078	51
Pom-21	0.1233	0.1238	45
Pom-22	0.0634	0.059	49
Pom-23	0.133	0.1346	39
Pom-24	0.1131	0.1076	43
Pom-25	0.0297	0.0351	35.5
Pom-26	0.0982	0.0896	47
Pom-27	0.1121	0.1052	43
Pom-28	0.1112	0.1009	44
Pom-29	0.1299	0.1178	41
Pom-30	0.0749	0.0704	45
Pom-31	0.0662	0.0613	42
Pom-32	0.1423	0.1384	41
Pom-33	0.1163	0.1161	45
Pom-34	0.0511	0.0569	39
Pom-35	0.0811	0.0715	51
Pom-36	0.105	0.0956	48
Pom-37	0.1402	0.1345	45
Pom-38	0.0945	0.0988	44
Pom-39	0.1282	0.1323	45
Pom-40	0.1024	0.0989	46
Pom-41	0.0845	0.0788	47
Pom-43	0.0925	0.0933	41
Pom-44	0.1402	0.1436	38
Pom-45	0.0669	0.0665	41
Pom-46	0.1074	0.1019	42
Pom-47	0.1463	0.1434	39
Pom-48	0.0662	0.0613	47

ID campione	Deamminazione in 5' (%)	amminazione in 3' (%)	imensione media dei frammenti
Pom-49	0.0371	0.0709	37
Pom-50	0.0978	0.0869	50
Pom-51	0.1178	0.1125	44
Pom-52	0.0574	0.0538	38
Pom-53	0.1017	0.1002	39
Pom-54	0.0624	0.0472	51
Pom-55	0.0929	0.0862	46
Pom-56	0.1027	0.0906	48
Pom-57	0.0419	0.0263	31
Pom-58	0.2092	0.2039	41
Pom-59	0.0732	0.0679	42
Pom-60	0.0141	0.0237	40
Pom-61	0.1027	0.0952	43
Pom-62	0.0624	0.0545	51
Pom-63	0.1267	0.1181	46
Pom-64	0.0725	0.0672	47
Pom-65	0.0227	0.0464	36
Pom-66	0.1001	0.0937	46
Pom-67	0.1403	0.1386	39
Pom-68	0.0987	0.1045	41
Pom-69	0.1008	0.0954	47
Pom-70	0.1367	0.1281	41
Pom-71	0.1365	0.1258	39
Pom-72	0.1175	0.116	42
Pom-73	0.1286	0.1232	48
Pom-74	0.1287	0.1081	41
Pom-75	0.0388	0.0238	30
Pom-76	0.1423	0.1418	44
Pom-77	0.0766	0.0746	49
Pom-78	0.139	0.1363	45
Pom-79	0.0787	0.081	46
Pom-80	0.1194	0.1183	45
Pom-81	0.2767	0.261	42
Pom-82	0.1033	0.104	43
Pom-83	0.1895	0.1544	39
Pom-84	0.0936	0.0935	40
Pom-85	0.1337	0.1269	39
Pom-86	0.1393	0.1402	41
Pom-87	0.0687	0.0689	46
Pom-88	0.0467	0.0462	43
Pom-89	0.023	0.0405	31
Pom-90	0.0988	0.0972	43
Pom-91	0.1075	0.1072	36
Pom-92	0.0179	0.0167	31
Pom-93	0.0077	0	31
Pom-96B	0.1365	0.1348	50
Pom-98	0.0712	0.0562	41
Pom-99	0.1587	0.1527	41
Pom-100	0.1343	0.1283	45
Pom-101	0.1235	0.1187	44

ID campione	Deamminazione in 5' (%)	Deamminazione in 3' (%)	Dimensione media dei frammenti
Pom-102	0.0745	0.072	47
Pom-103	0.1171	0.1111	46
Pom-104	0.1131	0.1069	50
Pom-105	0.0517	0.0401	51
Pom-106	0.12	0.1169	46
Pom-107	0.0803	0.0728	49
Pom-108	0.1354	0.1226	48
Pom-109	0.13	0.1225	40
Pom-110	0.1394	0.1339	41
Pom-111	0.0955	0.0905	47
Pom-112	0.1076	0.1007	46
Pom-113	0.1111	0.1003	49
Pom-114	0.1233	0.1164	46
Pom-115	0.0809	0.0782	45
Pom-116	0.1101	0.1019	45
Pom-117	0.1286	0.1251	42
Pom-118	0.0848	0.0727	51
Pom-119	0.101	0.0981	43
Pom-120	0.0668	0.0613	47
Pom-121	0.1341	0.1311	42
Pom-122	0.146	0.1448	45
Pom-123	0.0674	0.0603	51
Pom-124	0.099	0.0923	47
Pom-125	0.1099	0.1089	49
Pom-126	0.0712	0.0671	48
Pom-127	0.076	0.0655	40
Pom-128	0.0341	0.0307	47
Pom-129	0.0983	0.0956	46
Pom-130	0.0881	0.0901	45
Pom-131	0.0989	0.0959	44
Pom-132	0.0877	0.0886	48
Pom-133	0.085	0.082	46
Pom-134	0.1215	0.1166	45
Pom-135	0.1106	0.1043	40
Pom-136	0.115	0.1106	43
Pom-137	0.0678	0.0674	49
Pom-139	0.0824	0.0798	48
Pom-140	0.0296	0.0272	49
Pom-143	0.081	0.0711	44
Pom-144	0.1162	0.1077	42
Pom-145	0.0516	0.0422	42
Pom-146	0.1148	0.1013	39
Pom-147	0.0382	0.0324	40
Pom-158	0.0984	0.0983	45
Pom-159	0.1092	0.106	46
Pom-160	0.0958	0.0952	49
Pom-161	0.0675	0.0649	46
Pom-162	0.0994	0.0989	45
Pom-163	0.0868	0.0875	44
Pom-164	0.0418	0.0414	49

ID campione	Deaminazione in 5' (%)	Deaminazione in 3' (%)	Dimensione media dei frammenti
Pom-165A	0.0824	0.0693	45
Pom-165B	0.0661	0.0662	47
Pom-166	0.1368	0.1321	45

Tabella 4 Risultati ottenuti con MapDamage 2.0 sugli *shotgun light*.

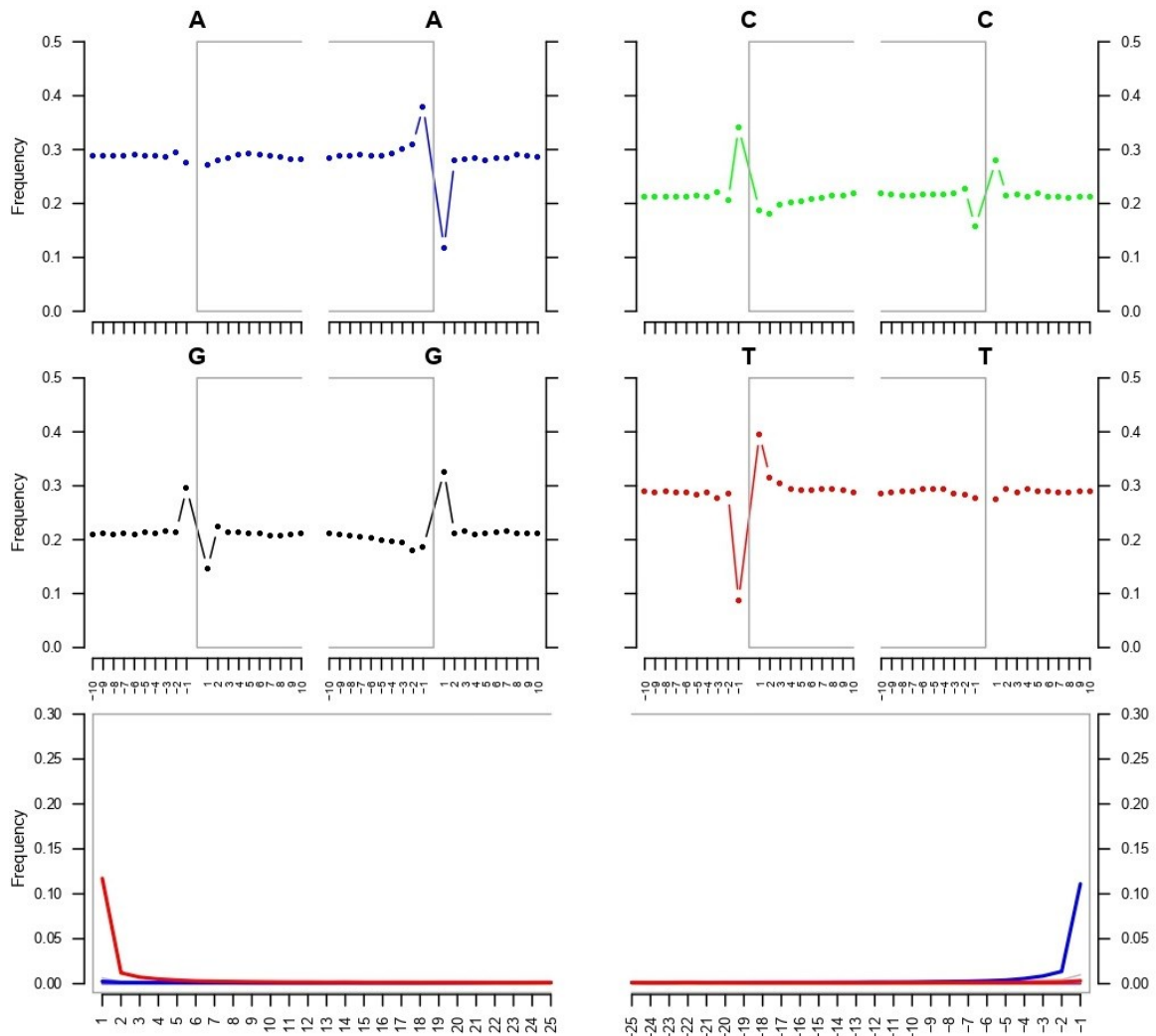


Figura 11 Plot ottenuto dal software MapDamage v2.0 rappresentante le transizioni rilevate alle estremità 5' e 3' delle molecole di DNA per il campione Pom-103. In alto nei quattro mini-plot è riportata la frequenza di ciascun nucleotide, mentre nel pannello in basso la percentuale di transizioni C->T (in rosso) e G->A (in blu).

I quattro grafici superiori nella *Figura 11* mostrano la distribuzione di frequenza dei 4 nucleotidi (A, C, G, T) lungo le prime 10 bp delle molecole di DNA sia in 5' che in 3'. All'interno delle parentesi quadre sono riportati i valori medi di frequenza riferiti alle *reads* del campione antico dalla posizione in 5'

(parentesi a sinistra) alla posizione in 3' (parentesi di destra), all'esterno invece i valori medi relativi alla sequenza di riferimento. Si osserva quindi un aumento della frequenza delle T all'estremità 5' delle molecole antiche e delle A all'estremità 3'. Nel grafico inferiore è riportata, per le prime 25 bp, in rosso la frequenza delle transizioni C->T in 5' e in blu quella G->A in 3'. Ancora una volta, coerentemente con quanto ci si aspetta dall'analisi dell'aDNA e dal trattamento *half-UDG* delle librerie che ripara le deaminazioni interne alla molecola lasciando gli ultimi due nucleotidi terminali danneggiati, la frequenza delle sostituzioni C->T e G->A aumenta rispettivamente in 5' e 3'.

4.1.1 Determinazione del sesso

Sui campioni sottoposti all'analisi preliminare per la valutazione dell'endogeno, in possesso di *coverage* medio diverso da 0, è stata eseguita l'analisi per l'identificazione del sesso su base molecolare come descritto nel *paragrafo 3.4.3*. I risultati ottenuti sono riportati in *Tabella 5*.

ID campioni	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assegnazione
Pom-1	1497783	38907	3231	0.083	0.0014	0.0803-0.0858	XY
Pom-2	480247	22561	79	0.0035	0.0004	0.0027-0.0043	XX
Pom-3	1468742	72507	227	0.0031	0.0002	0.0027-0.0035	XX
Pom-4/9	1829616	50461	4101	0.0813	0.0012	0.0789-0.0837	XY
Pom-5/6	1487789	72712	216	0.003	0.0002	0.0026-0.0034	XX
Pom-8	11359	499	0	0	0	0.0-0.0	consistent with XX
Pom-10.1	7073	369	3	0.0081	0.0047	-0.001-0.0173	consistent with XX but not XY
Pom-10	23326	1154	1	0.0009	0.0009	-0.0008-0.0026	XX
Pom-11	-	-	-	-	-	-	-
Pom-11.1	7563	358	7	0.0196	0.0073	0.0052-0.0339	consistent with XX but not XY
Pom-12	1486143	71623	276	0.0039	0.0002	0.0034-0.0043	XX
Pom-13	-	-	-	-	-	-	-
Pom-15	-	-	-	-	-	-	-
Pom-16A	293664	13609	88	0.0065	0.0007	0.0051-0.0078	XX
Pom-16B	238144	4260	18	0.0042	0.001	0.0023-0.0062	XX
Pom-17	-	-	-	-	-	-	-
Pom-18	-	-	-	-	-	-	-
Pom-19	208716	10368	40	0.0039	0.0006	0.0027-0.0051	XX
Pom-20	3166144	148623	508	0.0034	0.0002	0.0031-0.0037	XX
Pom-21	258090	7197	578	0.0803	0.0032	0.074-0.0866	consistent with XY but not XX
Pom-22	-	-	-	-	-	-	-
Pom-23	11546	538	14	0.026	0.0069	0.0126-0.0395	consistent with XX but not XY
Pom-24	1071115	50176	195	0.0039	0.0003	0.0033-0.0044	XX
Pom-25	-	-	-	-	-	-	-
Pom-26	2138653	58363	4747	0.0813	0.0011	0.0791-0.0836	XY
Pom-27	281362	7349	591	0.0804	0.0032	0.0742-0.0866	consistent with XY but not XX
Pom-28	1280897	60139	213	0.0035	0.0002	0.0031-0.004	XX
Pom-29	6667	245	1	0.0041	0.0041	-0.0039-0.0121	XX
Pom-30	2694582	125793	424	0.0034	0.0002	0.0031-0.0037	XX
Pom-31	71408	3359	17	0.0051	0.0012	0.0027-0.0075	XX
Pom-32	1075558	50633	210	0.0041	0.0003	0.0036-0.0047	XX
Pom-33	72633	3309	23	0.007	0.0014	0.0041-0.0098	XX
Pom-34	-	-	-	-	-	-	-
Pom-35	93449	2443	193	0.079	0.0055	0.0683-0.0897	consistent with XY but not XX
Pom-36	2565355	124176	368	0.003	0.0002	0.0027-0.0033	XX
Pom-37	1164378	31958	2555	0.0799	0.0015	0.077-0.0829	XY
Pom-38	12605	329	33	0.1003	0.0166	0.0678-0.1328	consistent with XY but not XX
Pom-39	34014	836	67	0.0801	0.0094	0.0617-0.0985	consistent with XY but not XX

ID campioni	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assegnazione
Pom-40	1890823	48933	4024	0.0822	0.0012	0.0798-0.0847	XY
Pom-41	2915042	76449	6337	0.0829	0.001	0.0809-0.0848	XY
Pom-43	27152	1273	13	0.0102	0.0028	0.0047-0.0157	XX
Pom-44	37804	1046	66	0.0631	0.0075	0.0484-0.0778	consistent with XY but not XX
Pom-45	314918	15006	46	0.0031	0.0005	0.0022-0.0039	XX
Pom-46	57695	1474	116	0.0787	0.007	0.065-0.0924	consistent with XY but not XX
Pom-47	59386	1748	127	0.0727	0.0062	0.0605-0.0848	consistent with XY but not XX
Pom-48	11009996	518073	1737	0.0034	0.0001	0.0032-0.0035	XX
Pom-49	-	-	-	-	-	-	-
Pom-50	843390	40364	137	0.0034	0.0003	0.0028-0.004	XX
Pom-51	183079	8346	42	0.005	0.0008	0.0035-0.0066	XX
Pom-52	-	-	-	-	-	-	-
Pom-53	629746	30648	98	0.0032	0.0003	0.0026-0.0038	XX
Pom-54	15322385	426258	36558	0.0858	0.0004	0.0849-0.0866	XY
Pom-55	303169	8114	635	0.0783	0.003	0.0724-0.0841	consistent with XY but not XX
Pom-56	7078821	171757	14190	0.0826	0.0007	0.0813-0.0839	XY
Pom-57	-	-	-	-	-	-	-
Pom-58	109649	3089	229	0.0741	0.0047	0.0649-0.0834	consistent with XY but not XX
Pom-59	-	-	-	-	-	-	-
Pom-60	-	-	-	-	-	-	-
Pom-61	1394360	38083	3125	0.0821	0.0014	0.0793-0.0848	XY
Pom-62	17097941	807477	2553	0.0032	0.0001	0.003-0.0033	XX
Pom-63	2291114	64075	5328	0.0832	0.0011	0.081-0.0853	XY
Pom-64	6640361	180927	14861	0.0821	0.0006	0.0809-0.0834	XY
Pom-65	-	-	-	-	-	-	-
Pom-66	6108488	292987	987	0.0034	0.0001	0.0032-0.0036	XX
Pom-67	446193	11770	957	0.0813	0.0025	0.0764-0.0862	XY
Pom-68	47371	2116	15	0.0071	0.0018	0.0035-0.0107	XX
Pom-69	939169	44199	129	0.0029	0.0003	0.0024-0.0034	XX
Pom-70	42084	1112	80	0.0719	0.0077	0.0568-0.0871	consistent with XY but not XX
Pom-71	21007	465	33	0.071	0.0119	0.0476-0.0943	consistent with XY but not XX
Pom-72	1349244	34202	2808	0.0821	0.0015	0.0792-0.085	XY
Pom-73	106785	2765	238	0.0861	0.0053	0.0756-0.0965	XY
Pom-74	6046	342	7	0.0205	0.0077	0.0055-0.0355	consistent with XX but not XY
Pom-75	-	-	-	-	-	-	-
Pom-76	113050	2997	252	0.0841	0.0051	0.0741-0.094	consistent with XY but not XX
Pom-77	2977461	77162	6331	0.082	0.001	0.0801-0.084	XY
Pom-78	3325941	154586	444	0.0029	0.0001	0.0026-0.0031	XX
Pom-79	1720237	46914	3823	0.0815	0.0013	0.079-0.084	XY
Pom-80	831124	37396	119	0.0032	0.0003	0.0026-0.0038	XX
Pom-81	-	-	-	-	-	-	-
Pom-82	30267	1481	4	0.0027	0.0013	0.0001-0.0053	XX
Pom-83	-	-	-	-	-	-	-
Pom-84	5792	306	1	0.0033	0.0033	-0.0031-0.0097	XX
Pom-85	316614	8360	698	0.0835	0.003	0.0776-0.0894	XY
Pom-86	36703	1084	78	0.072	0.0078	0.0566-0.0873	consistent with XY but not XX
Pom-87	2041850	97550	302	0.0031	0.0002	0.0027-0.0034	XX
Pom-88	2762788	133674	396	0.003	0.0001	0.0027-0.0033	XX
Pom-89	-	-	-	-	-	-	-
Pom-90	2485058	115974	392	0.0034	0.0002	0.003-0.0037	XX
Pom-91	-	-	-	-	-	-	-
Pom-92	-	-	-	-	-	-	-
Pom-93	-	-	-	-	-	-	-
Pom-96B	2551546	119298	80	0.0007	0.0001	0.0005-0.0008	XX
Pom-98	-	-	-	-	-	-	-

ID campioni	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assegnazione
Pom-99	610501	29056	88	0.003	0.0003	0.0024-0.0037	XX
Pom-100	1531729	40522	3300	0.0814	0.0014	0.0788-0.0841	XY
Pom-101	831102	21864	1777	0.0813	0.0018	0.0777-0.0849	XY
Pom-102	243631	11741	45	0.0038	0.0006	0.0027-0.005	XX
Pom-103	2638519	128562	394	0.0031	0.0002	0.0028-0.0034	XX
Pom-104	2823963	77395	6369	0.0823	0.001	0.0804-0.0842	XY
Pom-105	5283816	251093	812	0.0032	0.0001	0.003-0.0035	XX
Pom-106	1489629	39192	3226	0.0823	0.0014	0.0796-0.085	XY
Pom-107	5022464	133157	10869	0.0816	0.0008	0.0802-0.0831	XY
Pom-108	393006	19222	47	0.0024	0.0004	0.0017-0.0031	XX
Pom-109	-	-	-	-	-	-	-
Pom-110	1785076	86314	286	0.0033	0.0002	0.0029-0.0037	XX
Pom-111	743047	19510	1607	0.0824	0.002	0.0785-0.0862	XY
Pom-112	1573613	74197	250	0.0034	0.0002	0.003-0.0038	XX
Pom-113	4076683	193188	598	0.0031	0.0001	0.0028-0.0033	XX
Pom-114	247314	6553	499	0.0761	0.0033	0.0697-0.0826	consistent with XY but not XX
Pom-115	-	-	-	-	-	-	-
Pom-116	-	-	-	-	-	-	-
Pom-117	1018687	48763	148	0.003	0.0002	0.0025-0.0035	XX
Pom-118	6053789	161663	13481	0.0834	0.0007	0.082-0.0847	XY
Pom-119	-	-	-	-	-	-	-
Pom-120	-	-	-	-	-	-	-
Pom-121	876064	23480	1866	0.0795	0.0018	0.076-0.0829	XY
Pom-122	1554647	41733	3199	0.0767	0.0013	0.0741-0.0792	consistent with XY but not XX
Pom-123	5183246	247962	705	0.0028	0.0001	0.0026-0.0031	XX
Pom-124	11721614	572806	1426	0.0025	0.0001	0.0024-0.0026	XX
Pom-125	2873361	132602	106	0.0008	0.0001	0.0006-0.001	XX
Pom-126	2085721	98214	343	0.0035	0.0002	0.0031-0.0039	XX
Pom-127	2544574	67975	5129	0.0755	0.001	0.0735-0.0774	consistent with XY but not XX
Pom-128	3176361	140208	93	0.0007	0.0001	0.0005-0.0008	XX
Pom-129	1317797	35503	2744	0.0773	0.0014	0.0745-0.0801	consistent with XY but not XX
Pom-130	33646	1522	1	0.0007	0.0007	-0.0006-0.0019	XX
Pom-131	384652	18729	16	0.0009	0.0002	0.0004-0.0013	XX
Pom-132	1493994	40139	3246	0.0809	0.0014	0.0782-0.0835	XY
Pom-133	97495	2655	208	0.0783	0.0052	0.0681-0.0886	consistent with XY but not XX
Pom-134	88239	2414	166	0.0688	0.0052	0.0587-0.0789	consistent with XY but not XX
Pom-135	133448	6411	6	0.0009	0.0004	0.0002-0.0017	XX
Pom-136	618208	16882	1343	0.0796	0.0021	0.0755-0.0836	XY
Pom-137	316806	8475	657	0.0775	0.0029	0.0718-0.0832	consistent with XY but not XX
Pom-139	444589	12206	907	0.0743	0.0024	0.0697-0.079	consistent with XY but not XX
Pom-140	596625	16507	1223	0.0741	0.002	0.0701-0.0781	consistent with XY but not XX
Pom-143	183621	4899	372	0.0759	0.0038	0.0685-0.0834	consistent with XY but not XX
Pom-144	381735	18197	11	0.0006	0.0002	0.0002-0.001	XX
Pom-145	107821	3010	192	0.0638	0.0045	0.0551-0.0725	NA
Pom-146	182230	8747	2	0.0002	0.0002	-0.0001-0.0005	XX
Pom-147	7089	339	5	0.0147	0.0065	0.0019-0.0276	consistent with XX but not XY
Pom-158	1212762	32551	2484	0.0763	0.0015	0.0734-0.0792	consistent with XY but not XX
Pom-159	1564109	74506	38	0.0005	0.0001	0.0003-0.0007	XX
Pom-160	3537450	94960	7222	0.0761	0.0009	0.0744-0.0777	consistent with XY but not XX
Pom-161	4735259	229255	125	0.0005	0	0.0004-0.0006	XX
Pom-162	3831389	185296	79	0.0004	0	0.0003-0.0005	XX
Pom-163	169233	8401	5	0.0006	0.0003	0.0001-0.0011	XX
Pom-164	66184	3228	0	0	0	0.0-0.0	consistent with XX
Pom-165A	102128	2738	182	0.0665	0.0048	0.0571-0.0758	consistent with XY but not XX
Pom-165B	400015	10987	834	0.0759	0.0025	0.071-0.0809	consistent with XY but not XX
Pom-166	384314	9062	684	0.0755	0.0028	0.07-0.0809	consistent with XY but not XX

Tabella 5 Risultati relativi al sesso determinato su base molecolare. Gli individui segnati in rosso sono coloro il cui sesso è stato determinato con incertezza a causa del basso *coverage*.

In *Tabella 5* sono riportati, per ciascun individuo, il numero di *reads* mappanti (Nseq), le *reads* mappanti su entrambi i cromosomi sessuali (NchrY+NchrX), quelle sul cromosoma Y (NchrY), il rapporto tra NchrY e NchrY+NchrX (Ry), l'errore standard (SE) e l'intervallo di confidenza al 95% (95% CI).

Per la maggior parte degli individui è stato possibile determinare il sesso contro ogni ragionevole dubbio, mentre in alcuni casi il numero di *reads* non era sufficiente affinché l'assegnazione fosse certa (righe segnate in rosso in *Tabella 5*).

4.1.2 Genotyping e analisi delle componenti principali (PCA)

Il genoma degli individui il cui *coverage* medio, in seguito al primo sequenziamento *shotgun* esplorativo, risulta diverso da 0 è stato genotipizzato per due *SNPs-array* tramite un approccio basato sulla chiamata pseudo-aploide (come descritto nel *paragrafo 3.4.6*). I risultati relativi alla copertura degli SNPs per ciascun individuo sono riportati in *Tabella 6*. I soli campioni genotipizzati sul pannello *Human Origins* (HO) e con almeno 10000 SNPs tipizzati sono stati utilizzati per l'analisi delle componenti principali (PCA).

ID campioni	SNPs coperti HO	SNPs coperti 1240K
Pom-1	20757	37736
Pom-2	6454	11943
Pom-3	19889	37719
Pom-4 /Pom-9	23879	44456
Pom-5 /Pom-6	19043	35557
Pom-7	1583	2940
Pom-8	64	112
Pom-10	114	260
Pom-10.1	71	136
Pom-11	-	-
Pom-11.1	80	154
Pom-12	20241	37659
Pom-13	-	-
Pom-15	-	-
Pom-16A	2019	4105
Pom-16B	693	1418
Pom-17	-	-
Pom-18	-	-
Pom-19	3026	5824
Pom-20	41189	75973
Pom-21	15411	29130
Pom-22	-	-
Pom-23	103	226
Pom-24	12286	23074
Pom-25	-	-
Pom-26	26338	49398
Pom-27	3463	6267

ID campioni	SNPs coperti HO	SNPs coperti 1240K
Pom-28	15368	28639
Pom-29	66	110
Pom-30	32509	60017
Pom-31	846	1588
Pom-32	12348	22983
Pom-33	949	1704
Pom-34	-	-
Pom-35	1324	2381
Pom-36	30229	57613
Pom-37	13296	25336
Pom-38	158	294
Pom-39	433	778
Pom-40	24075	44535
Pom-41	37526	68780
Pom-43	338	599
Pom-44	420	759
Pom-45	3873	7039
Pom-46	680	1268
Pom-47	610	1172
Pom-48	123331	229235
Pom-49	-	-
Pom-50	10353	19913
Pom-51	2290	4138
Pom-52	-	-
Pom-53	6808	12902
Pom-54	154115	300119
Pom-55	3936	7222

ID campioni	SNPs coperti HO	SNPs coperti 1240K
Pom-56	84266	154448
Pom-57	-	-
Pom-58	1199	2308
Pom-59	-	-
Pom-60	-	-
Pom-61	17045	31328
Pom-62	185054	345389
Pom-63	27070	51542
Pom-64	79922	148941
Pom-65	-	-
Pom-66	68256	129886
Pom-67	5031	9482
Pom-68	576	1031
Pom-69	11267	21465
Pom-70	586	1046
Pom-71	259	456
Pom-72	16337	29907
Pom-73	1539	2858
Pom-74	70	134
Pom-75	-	-
Pom-76	1402	2621
Pom-77	39657	73557
Pom-78	40551	75362
Pom-79	21704	40584
Pom-80	10770	19908
Pom-81	-	-
Pom-82	338	655
Pom-83	-	-
Pom-84	50	99
Pom-85	3573	6578
Pom-86	421	815
Pom-87	25973	48456
Pom-88	32773	61752
Pom-89	-	-
Pom-90	27720	52006
Pom-91	-	-
Pom-92	-	-
Pom-93	-	-
Pom-96B	33746	62999
Pom-98	-	-
Pom-99	6849	12976
Pom-100	18811	34635
Pom-101	9626	18109
Pom-102	2877	5469
Pom-103	28752	55201
Pom-104	33882	64291
Pom-105	66460	123907
Pom-106	18571	34157

ID campioni	SNPs coperti HO	SNPs coperti 1240K
Pom-107	61128	114030
Pom-108	4566	8843
Pom-109	-	-
Pom-110	19039	36246
Pom-111	9583	17720
Pom-112	18450	34632
Pom-113	47890	90294
Pom-114	3153	5812
Pom-115	-	-
Pom-116	-	-
Pom-117	11235	21482
Pom-118	76695	142131
Pom-119	-	-
Pom-120	-	-
Pom-121	9885	18648
Pom-122	18833	35740
Pom-123	62475	117258
Pom-124	124601	237313
Pom-125	38523	71540
Pom-126	26127	48281
Pom-127	26487	50417
Pom-128	37667	70419
Pom-129	17006	31771
Pom-130	435	797
Pom-131	4507	8613
Pom-132	20585	38267
Pom-133	1293	2442
Pom-134	1101	2039
Pom-135	1433	2744
Pom-136	7250	13606
Pom-137	4343	8068
Pom-139	6020	11579
Pom-140	8223	15674
Pom-143	2158	4037
Pom-144	4054	7840
Pom-145	1238	2370
Pom-146	1845	3555
Pom-147	58	127
Pom-158	14997	27863
Pom-159	19884	37644
Pom-160	47806	89706
Pom-161	59309	111915
Pom-162	43653	84419
Pom-163	1901	3715
Pom-164	920	1741
Pom-165A	1254	2376
Pom-165B	5338	10133
Pom-166	4950	8984

Tabella 6 SNPs coperti in seguito alla genotipizzazione effettuata sia sul pannello *Human Origins* (HO 597573 SNPs totali) che sul 1240K (1233013 SNPs totali). In rosso sono indicati gli individui con meno di 10000 SNPs coperti sul pannello HO e non usati per la PCA.

La PCA consente una prima visualizzazione della variabilità genetica che caratterizza la popolazione esaminata.

In *Figura 12* è riportata la PCA in cui in grigio sono indicate le attuali popolazioni dell'Eurasia occidentale e del Nord Africa utilizzate per la costruzione dello spazio di variabilità genetica. Sono state utilizzate la PC1 e la PC2 (rappresentante rispettivamente lo 0.75% e lo 0.36% della varianza totale) per proiettare in grafico la variabilità degli individui antichi di confronto già presenti in letteratura e rappresentati con dei cerchi colorati (individui all'interno del dataset HO v44.3 più ulteriori campioni rilasciati in seguito alla pubblicazione del dataset: Saupe et al., 2021; Posth et al., 2021; Aneli et al., 2022; Antonio et al., 2023 bioRxiv) e gli individui ritrovati a Pompei indicati con dei rombi.

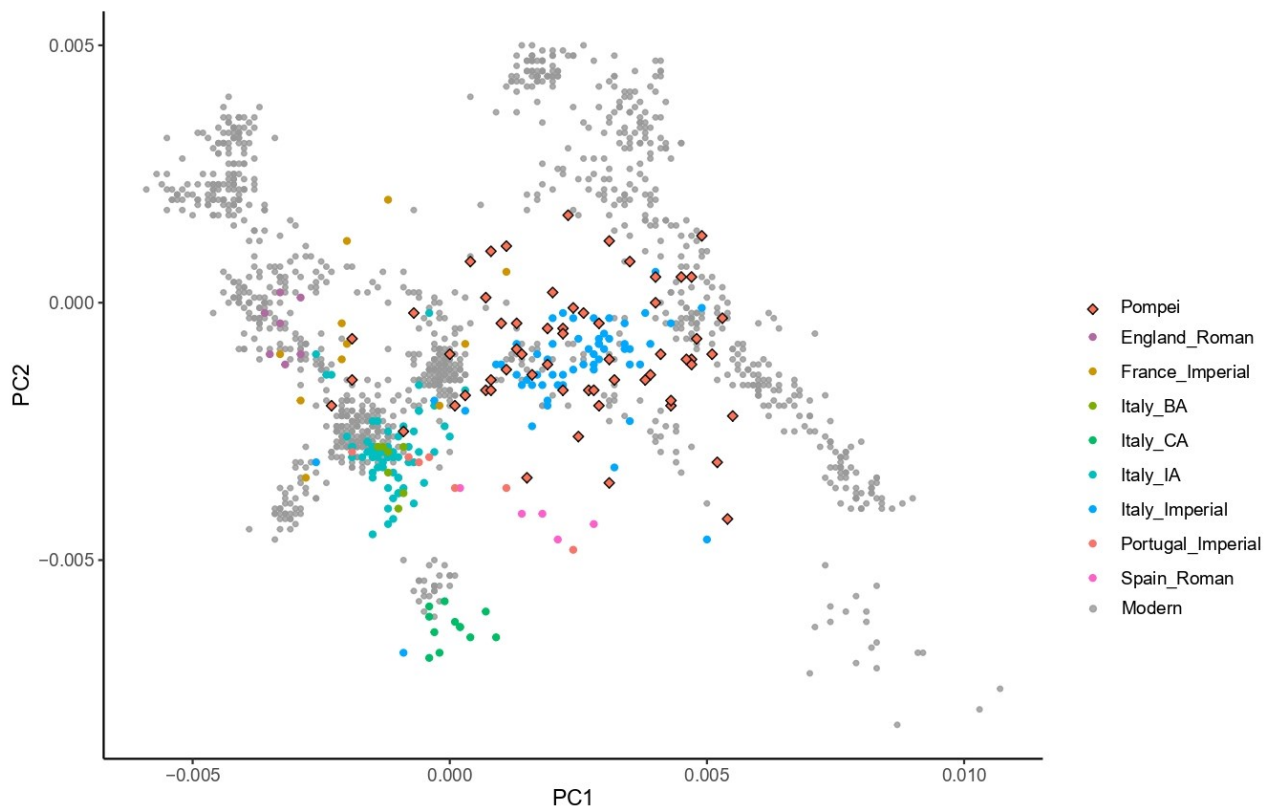


Figura 12 Analisi delle componenti principali realizzata a partire dalla variabilità genomica di 61 popolazioni dell'Eurasia occidentale genotipizzate per il pannello HO in cui sono stati proiettati i 62 individui ritrovati a Pompei e i campioni antichi di confronto.

È evidente l'elevatissima variabilità genetica dei campioni oggetto di questo studio. Infatti, essi sono distribuiti su tutti i clini di variabilità delle popolazioni moderne usate per la costruzione della PCA. La maggior parte degli individui a Pompei è sovrapposta alle attuali popolazioni del Mediterraneo centro-orientale, ma con delle eccezioni significative. Molti di essi, infatti, si discostano dal *cluster* principale di variabilità localizzandosi sulle attuali popolazioni dell'Europa occidentale e addirittura allontanandosi dalla variabilità Eurasiatica (verso le popolazioni del Nord Africa). Non solo gli individui a Pompei risultano diversi gli uni dagli altri, ma mostrano una variabilità genetica maggiore rispetto alla variabilità ritrovata in individui Imperiali già pubblicati e campionati nella città di Roma e nei dintorni di essa (*Italy_Imperial* in legenda). Il confronto con le popolazioni antiche presenti in letteratura, per gli individui di Pompei che fuoriescono dal cluster di variabilità prettamente Imperiale e localizzati sulle attuali popolazioni dell'Europa dell'Ovest, mostra una vicinanza di quest'ultimi con gli individui delle province Romane inglesi (*England_Roman*), con le popolazioni francesi del periodo Imperiale (*France_Imperial*) e con le popolazioni dell'IA Italiano (*Italy_IA*). Infine, gli individui la cui variabilità sembra vicina a quelle delle attuali popolazioni del Nord Africa si localizzano in prossimità delle popolazioni Imperiali della penisola Iberica (*Spain_Roman* e *Portugal_Imperial*) in cui è già stata identificata, a livello genomico, una componente Africana (Olalde et al., 2019; Antonio et al., 2023 bioRxiv).

4.2 Shotgun profondo

In seguito al primo sequenziamento *shotgun* esplorativo sono stati selezionati 12 individui tra quelli la cui percentuale di DNA endogeno superava il 20%. Inoltre, la scelta delle librerie da sottoporre al sequenziamento *deep shotgun* è stata effettuata tenendo conto della localizzazione degli individui in PCA (Figura 13). Sono stati selezionati gli individui la cui variabilità genetica rientra all'interno della variabilità già identificata in Italia nel periodo Imperiale (Antonio et al., 2019; Posth et al., 2021) e altri che si discostano da essa in modo da indagare più nel dettaglio la grande diversità che si trovava a Pompei prima che la città venisse distrutta. Gli individui scelti per questo secondo sequenziamento sono: Pom-20, Pom-41, Pom-77, Pom-78, Pom-88, Pom-104, Pom-105, Pom-107, Pom-110, Pom-112, Pom-123 e Pom-124.

Il processamento dei dati seguendo la *pipeline* descritta nel paragrafo 3.4.2 ha prodotto dei genomi completi le cui principali statistiche sono riportate in Tabella 7.

ID campioni	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Coverage 1X (%)
Pom-20	315355754	50766004	56.973	0.8706	49.16
Pom-41	288033346	48897952	57.652	0.7845	45.64
Pom-77	363934738	88954783	94.274	1.5575	64.03
Pom-78	1370152184	76826608	28.391	1.2313	58.82
Pom-88	301058704	68432735	80.422	1.1035	58.4
Pom-104	951325962	115240445	32.039	1.9348	72.58
Pom-105	156609976	50167463	95.14	1.0317	55.92
Pom-107	291849948	86132781	93.065	1.4587	64.85
Pom-110	741284158	79851240	28.647	1.1522	59.36
Pom-112	498860104	55159152	29.961	0.8988	52.69
Pom-123	301240532	68901177	63.048	1.2264	61.45
Pom-124	485597608	143060275	94.769	2.3355	75.95

Tabella 7 Principali statistiche relative al sequenziamento *deep shotgun* delle 12 librerie selezionate.

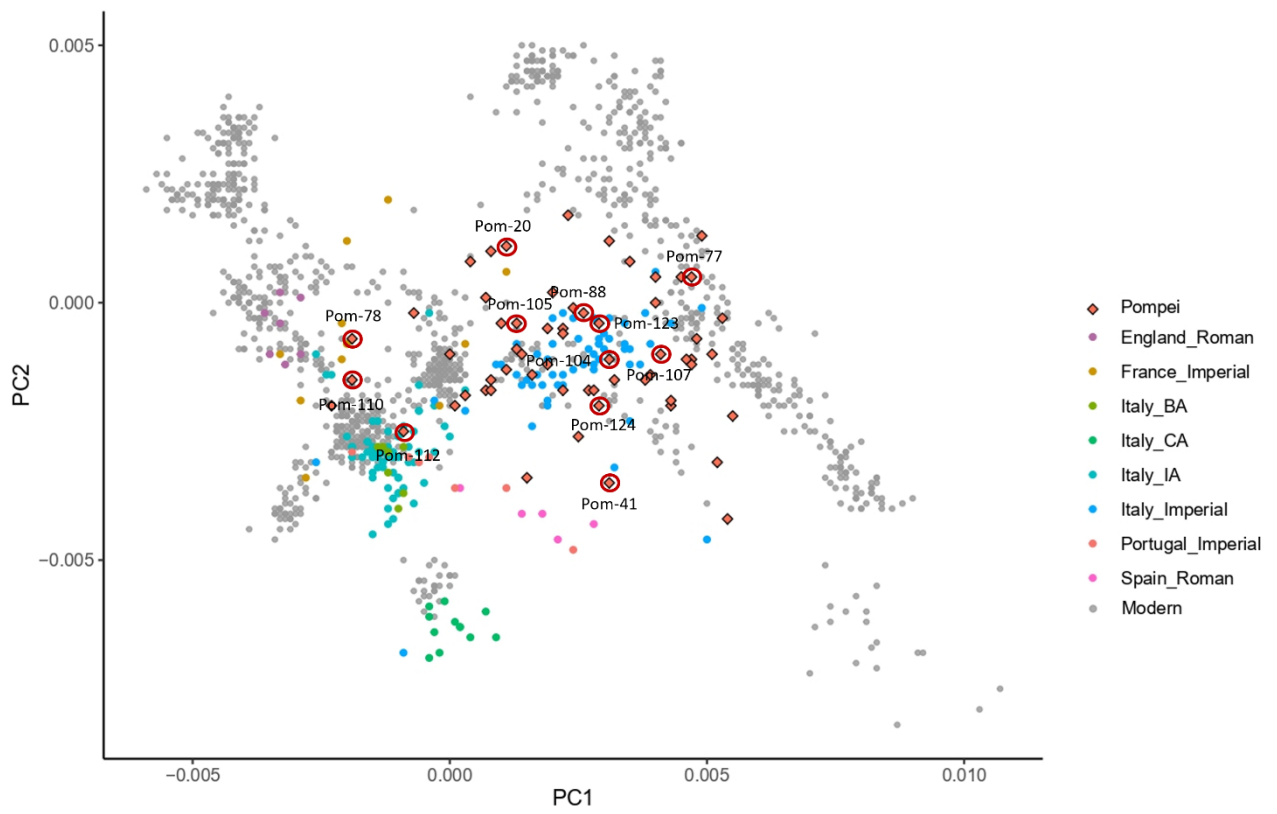


Figura 13 Analisi delle Componenti Principali come in *Figura 12* in cui sono evidenziati i 12 individui selezionati per il sequenziamento *deep shotgun*.

È stato raggiunto un *coverage* medio di 1x, ottenendo in alcuni casi addirittura circa il 76% del genoma coperto 1 volta. Come è possibile osservare in *Tabella 8* tutti i campioni mostrano i danni tipici delle molecole di aDNA successivamente al trattamento *half-UDG* delle librerie.

ID campione	Deamminazione in 5'	Deamminazione in 3'	Dimensione media dei frammenti
Pom-20	0.0958	0.0936	54
Pom-41	0.0856	0.0834	50
Pom-77	0.0753	0.0736	55
Pom-78	0.1373	0.1344	50
Pom-88	0.0460	0.0452	50
Pom-104	0.116	0.1162	53
Pom-105	0.0554	0.0551	64
Pom-107	0.0818	0.0818	53
Pom-110	0.1406	0.1398	45
Pom-112	0.1131	0.1109	51
Pom-123	0.0703	0.0699	56
Pom-124	0.1016	0.1020	51

Tabella 8 Valori di deamminazione per i 12 individui *deep shotgun* ottenuti con MapDamage v2.2.1.

4.2.1 Determinazione del sesso

L'applicazione del metodo di Skoglund et al. (2013) sui 12 individui in esame al fine di determinare il sesso su base molecolare ha confermato quanto ottenuto con i dati esplorativi. I risultati con le nuove coperture sono riportati in *Tabella 9*.

ID campioni	NchrY+NchrX	NchrY	Ry	SE	95% IC	Assegnazione
Pom-20	2339300	1604	0.0007	0	0.0007-0.0007	XX
Pom-41	1266076	97948	0.0774	0.0002	0.0769-0.0778	XY
Pom-77	2288504	177368	0.0775	0.0002	0.0772-0.0779	XY
Pom-78	3543296	1886	0.0005	0	0.0005-0.0006	XX
Pom-88	3307350	1782	0.0005	0	0.0005-0.0006	XX
Pom-104	3148552	244359	0.0776	0.0002	0.0773-0.0779	XY
Pom-105	2401663	1579	0.0007	0	0.0006-0.0007	XX
Pom-107	2298589	176243	0.0767	0.0002	0.0763-0.077	XY
Pom-110	3869841	2132	0.0006	0	0.0005-0.0006	XX
Pom-112	2620534	1738	0.0007	0	0.0006-0.0007	XX
Pom-123	3330307	1861	0.0006	0	0.0005-0.0006	XX
Pom-124	7057296	3003	0.0004	0	0.0004-0.0004	XX

Tabella 9 Risultati relativi al sequenziamento *deep shotgun* per la determinazione del sesso. Nseq numero di reads mappanti, NchrY+NchrX reads mappanti su entrambi i cromosomi sessuali, NchrY reads mappanti sul cromosoma Y, Ry rapporto tra NchrY e NchrY+NchrX, SE errore standard, 95% CI intervallo di confidenza al 95%.

4.2.2 Stima delle contaminazioni

Per i quattro individui di sesso maschile è stato possibile verificare la contaminazione nucleare tramite il metodo descritto nel *paragrafo 3.4.4*. Per ciascun individuo coinvolto, le stime di contaminazione ottenute sono riportate nella *Tabella 10* accompagnate dal proprio errore standard e dal numero di SNPs coinvolto nell'analisi. Tutti i campioni mostrano un livello di contaminazione trascurabile poiché inferiore allo 0.5%.

ID campione	X contamination ± SE	SNPs X-chr
Pom-41	0.0032±0.0012	3327
Pom-77	0.0041±0.0008	8924
Pom-104	0.0053±0.0008	13666
Pom-107	0.0048±0.001	9055

Tabella 10 Stima della contaminazione nucleare negli individui di sesso maschile

Inoltre, per tutti gli individui, sia maschi che femmine, è stato stimato il valore di contaminazione a livello mitocondriale come descritto nel *paragrafo 3.4.4*. Anche in questo caso i valori di contaminazione risultano trascurabili come mostrato in *Tabella 11* nella quale sono anche riportate il numero di *reads* mappanti sul mitocondrio, il *coverage* medio sul DNA mitocondriale e il rapporto *mt/nuclear ratio*. Proprio valutando quest'ultimo parametro (< 200) e considerando che il materiale di partenza per l'analisi degli individui consisteva, nella maggior parte dei casi, in rocche petrose, è possibile estendere l'assenza di contaminazione nucleare anche negli individui di sesso femminile come suggerito da Furtwängler et al. (2018).

Sample ID	Reads on MT	Coverage MT	Contaminazione mtDNA	mt/nuc Ratio
Pom-20	22662	72.42	0.01 (0-0.2)	83.22
Pom-41	26646	81.09	0.01 (0-0.2)	103.43
Pom-77	31060	100.73	0.01 (0-0.2)	64.7
Pom-78	35908	109.69	0.01 (0-0.2)	89.3
Pom-88	28472	84.64	0.01 (0-0.2)	76.73
Pom-104	51369	165.9	0.01 (0-0.2)	85.78
Pom-105	16187	60.61	0.01 (0-0.2)	58.76
Pom-107	26077	83.85	0.01 (0-0.2)	57.5
Pom-110	38998	108.41	0.01 (0-0.2)	94.14
Pom-112	38031	116.74	0.01 (0-0.2)	129.97
Pom-123	27000	87.47	0.01 (0-0.2)	71.35
Pom-124	34127	104.24	0.01 (0-0.2)	44.64

Tabella 11 Statistiche sul DNA mitocondriale e stima della contaminazione.

4.2.3 Identificazione degli aplogruppi uniparentali

L'assegnazione degli aplogruppi uniparentali (mtDNA e chrY) si è conclusa con successo per tutti gli individui in esame. In *Tabella 12* sono riportati, per i 12 individui analizzati, gli aplogruppi mitocondriali con i relativi valori di qualità e gli aplogruppi del chrY con il numero di marcatori validi usati per l'identificazione.

ID campione	Aplogruppo mtDNA	Qualità	Aplogruppo chrY	Marcatori validi
Pom-20	H	0.82	-	-
Pom-41	J1c1	0.92	T1a1a	5438
Pom-77	H14a	0.95	J2a2a1a~	14382
Pom-78	HV0+195	0.97	-	-
Pom-88	H1e1b1	0.9	-	-
Pom-104	J1d1a	0.96	J2a1a1a2b2a1a1a~	19363
Pom-105	H1+16189	0.91	-	-
Pom-107	H14b	0.85	J2a1a1b3	13352
Pom-110	U5a1+@16192	0.97	-	-
Pom-112	H13b1+200	0.91	-	-
Pom-123	HV1a3a	0.93	-	-
Pom-124	H13a2b3	0.92	-	-

Tabella 12 Aplogruppi mitocondriali e del cromosoma Y identificati nei 12 individui *deep shotgun*.

Tre dei quattro individui di sesso maschile possiedono l'aplogruppo del chrY appartenente al macro aplogruppo J2, già attestato nella penisola Iberica in periodo Imperiale (Olalde et al., 2019) e associato alla presenza dei Fenici (Zalloua et al., 2008). L'aplogruppo T è invece ampiamente distribuito in Europa, Medio Oriente, India e corno d'Africa e sebbene l'aplogruppo Y di Pom-41 rientri in questa variabilità è anche l'unica assegnazione supportata da un numero minore di marcatori validi. Gli aplogruppi del mtDNA rientrano invece all'interno della variabilità mitocondriale già osservata in Europa. I macro gruppi HV, H e J sono quelli più diffusi nel continente, comparsi per la prima volta in seguito all'ultima glaciazione e aumentati in frequenza successivamente al Neolitico (Posth et al., 2016; Modi et al., 2021). Anche l'aplogruppo U5, comparso nel tardo Paleolitico/inizi del Mesolitico e quindi arrivato con i gruppi di HGs dai rifugi glaciali, risulta essere molto comune in Europa (Posth et al., 2016; Modi et al., 2021).

4.2.4 Genotyping e Analisi delle Componenti Principali (PCA)

I 12 individui sono stati genotipizzati, ancora una volta, per lo *SNPs-array* HO con chiamata pseudo-aploide. Grazie al dato *deep shotgun* il numero di SNPs genotipizzato è maggiore rispetto a quello riportato, per gli stessi individui, in tabella *Tabella 6*. I nuovi dati di copertura sono riportati in *Tabella 13*.

ID campione	SNPs coperti HO
Pom-20	402969
Pom-41	388860
Pom-77	503104
Pom-78	469156
Pom-88	458467
Pom-104	538985
Pom-105	441706
Pom-107	506519
Pom-110	454986
Pom-112	407613
Pom-123	475468
Pom-124	562097

Tabella 13 Numero di SNPs genotipizzati per il pannello HO (597573) negli individui *deep shotgun*

L'analisi delle componenti principali (PCA), costruita con le stesse popolazioni moderne usate per la PCA con il dato *light shotgun* e confrontando i 12 individui di Pompei con le stesse popolazioni antiche disponibili in letteratura, ha portato ai risultati visibili nella *Figura 14*.

Sebbene la disposizione in generale della variabilità genetica dei 12 individui studiati confermi quella ottenuta con i dati di *shotgun* esplorativo (*Figura 13*), si possono però notare delle differenze. In *Figura 14* gli individui sovrapposti alla variabilità delle attuali popolazioni del Mediterraneo centro-occidentale, insieme ai contemporanei della capitale dell'Impero e nei dintorni di essa, tendono ad avvicinarsi gli uni agli altri. A conferma di una grande variabilità genetica all'interno dell'antica città di Pompei vi sono 6 individui che continuano a rimanere al di fuori del *cluster* prettamente Imperiale con un individuo che, addirittura, si avvicina ulteriormente alle popolazioni Imperiali della penisola Iberica in cui è attestata la presenza della componente nord Africana (Olalde et al., 2019; Antonio et al., 2023 bioRxiv).

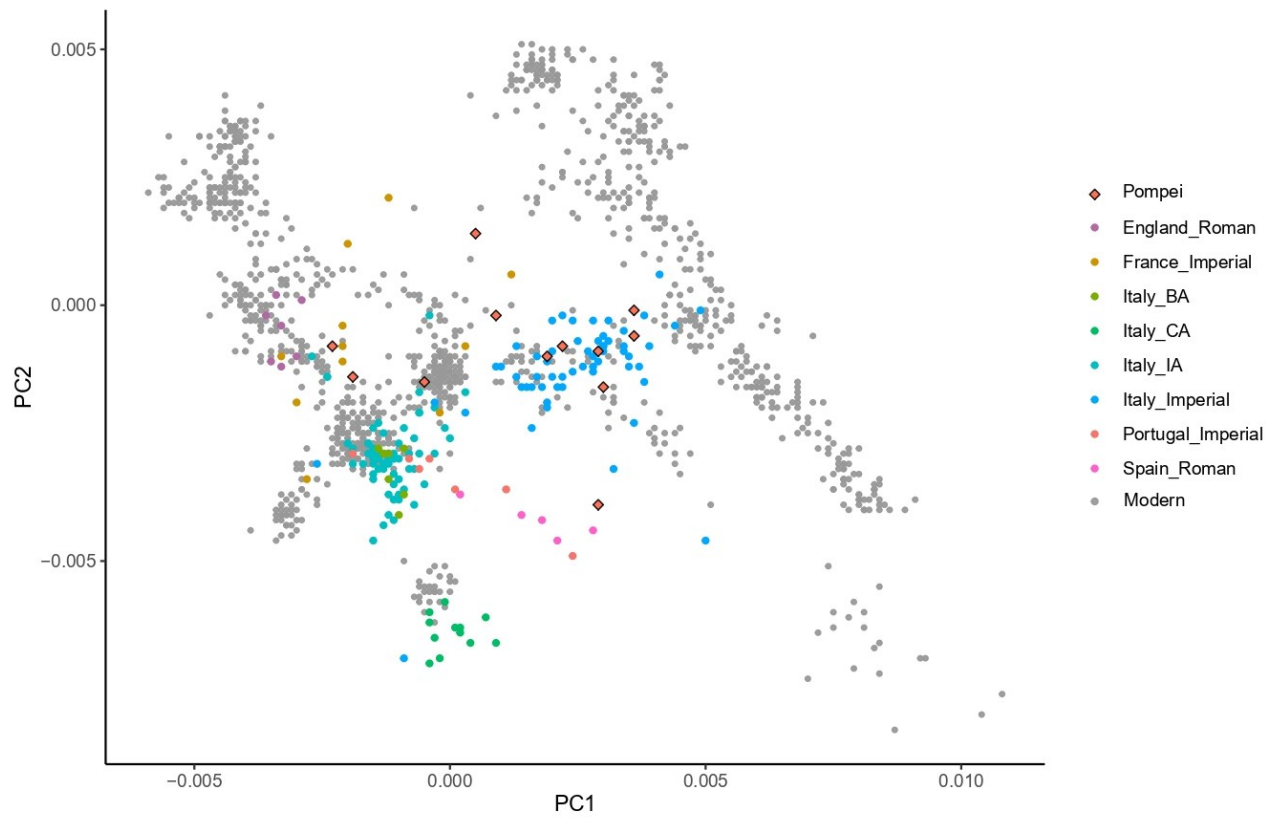


Figura 14 Analisi delle Componenti Principali (PCA) realizzata a partire dalla variabilità genomica di 61 popolazioni dell'Eurasia occidentale e dell'Africa del nord genotizzate per il pannello HO in cui sono stati proiettati i 12 individui *deep shotgun* e i campioni antichi di confronto.

4.2.5 Approccio *Whole Genome*

La statistica relativa alla copertura dei siti genomici letti da almeno 1, 3 e 5 *reads* nei 12 individui *deep shotgun* è riportata in *Tabella 14*.

ID Campione	Coverage (bp)		
	≥1	≥3	≥5
Pom-20	1530616397	288297968	36582956
Pom-41	1424137020	245093341	30198522
Pom-77	1998914455	732685586	210577990
Pom-78	1836660074	522318521	103761361
Pom-88	1824625685	414155900	51136264
Pom-104	2270155220	1012763099	284961953
Pom-105	1743362699	367293224	44132592
Pom-107	2026821591	677301025	142291453
Pom-110	1854404896	450384623	61905186
Pom-112	1642838290	272796282	24977429
Pom-123	1918552939	498431068	70334864
Pom-124	2377500451	1294420641	458082297

Tabella 14 Numero di nucleotidi letti almeno 1, 3 e 5 volte negli individui con *coverage* medio 1x.

È stato calcolato il numero di nucleotidi condivisi tra tutti e 12 i genomi in esame (100% degli individui) e in percentuali sempre più basse al fine di massimizzare il numero di posizioni coperte e condivise nel più grande numero possibile di individui. Inoltre, alle regioni del genoma condivise nelle varie percentuali di individui sono state applicate delle maschere, come descritto nel *paragrafo 3.4.8*, al fine di condurre le successive analisi solo a livello delle regioni neutrali, evitando quindi bias causati dall'adattamento ambientale. Il numero di paia basi in comune tra le varie percentuali di individui prima e dopo l'applicazione delle maschere sulle regioni autosomiche del genoma è riportato in *Tabella 15*.

Posizioni (bp) in varie % di individui	Coverage (bp)					
	≥1	≥3	≥5	≥1	≥3	≥5
	Senza maschere			Con maschere		
100% (12)	189948885	284487	41484	34701375	39885	-
90% (11)	553339875	1609193	78846	103616249	239023	-
80% (10)	977786808	6979737	130470	186592871	1100020	-
70% (8)	1716405601	62185602	437603	335531558	10616675	-
60% (7)	1980393317	139442504	1166927	389712885	24494957	-
50% (6)	2176490437	271941528	3842228	430002799	49039492	-
40% (5)	2315549317	474748885	13091769	458191248	87721669	-
30% (4)	2412368448	759088501	42032952	477212243	143452147	-
20% (2)	2530774922	1574920999	346262492	498275970	309289945	-

Tabella 15 Coperture a livello dell'intero genoma e nelle sole regioni neutrali condivise da varie percentuali di individui.

Le successive analisi (PCA e Admixture) hanno visto il calcolo delle *genotype likelihood* (GL) limitato alle posizioni autosomiche neutrali e polimorfiche (filtrate con il parametro *-SNP_pval 1e-6*) coperte

almeno una volta e in comune nell'80% degli individui. L'utilizzo delle sole posizioni condivise tra gli individui in esame fa sì che le analisi a valle risultino più robuste.

4.2.5.1 PCAnsd

L'Analisi delle Componenti Principali (PCA) è stata realizzata partendo dalle GL di 783840 posizioni (risultato del filtraggio a partire dal dato in *Tabella 15*), e ha visto il solo confronto dei 12 individui di Pompei con popolazioni antiche già presenti in letteratura i cui genomi sono stati selezionati secondo i criteri già descritti nel *paragrafo 3.4.8.1*. Nonostante si sia scelto di utilizzare le sole posizioni neutrali e condivise tra gli individui di Pompei, grazie all'analisi *whole-genome* è stato raggiunto un numero di SNPs maggiore di quello del pannello HO (597573 siti), tipicamente impiegato per lo stesso tipo di analisi con dati pseudo-aploidi. La distribuzione di variabilità genomica sulla PC1 e PC2 (rappresentate rispettivamente il 2.61% e lo 0.94% della varianza totale) ha restituito i risultati mostrati nella *Figura 15*.

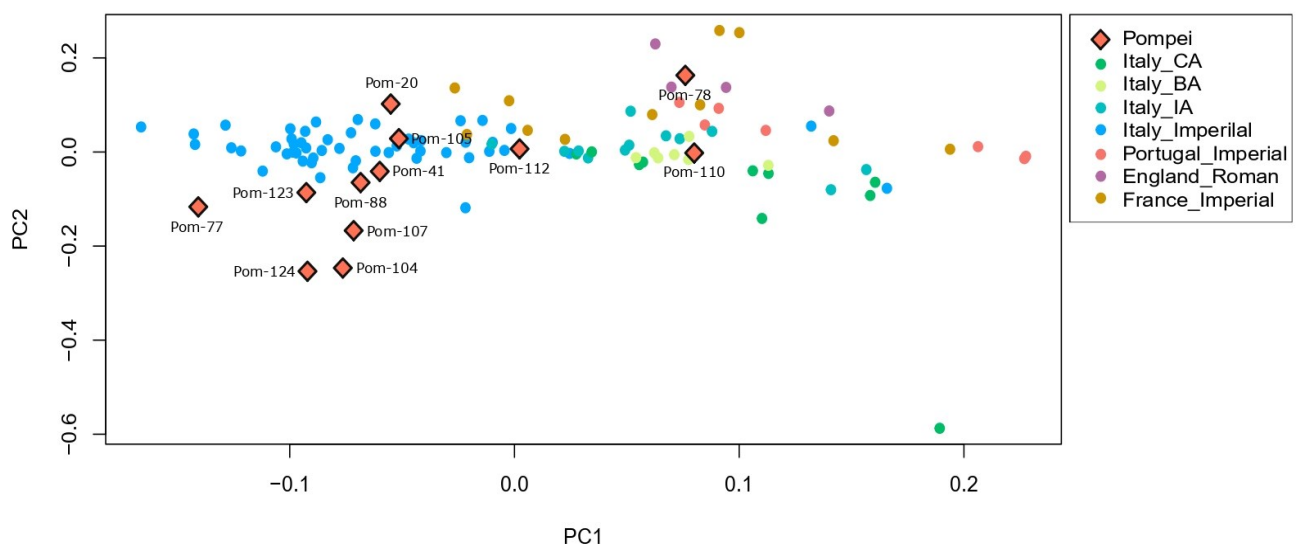


Figura 15 Analisi delle Componenti Principali (PCA) costruita a partire dalle GL calcolate sugli individui di Pompei e sulle popolazioni antiche già sequenziate con approccio di tipo *shotgun*.

In *Figura 15* la distribuzione delle popolazioni antiche di confronto risulta coerente con quanto già emerso dall'analisi dei dati pseudo-aploidi delle stesse. Il confronto di quest'ultime popolazioni antiche con gli individui ritrovati a Pompei conduce alle seguenti considerazioni:

1. La maggior parte degli individui di Pompei si sovrappone alle popolazioni Imperiali di Roma e dintorni (*Italy_Imperial*), anche se risulta una maggiore dispersione in PCA dei primi rispetto

alla nuvola di punti più compatta rappresentante gli Imperiali Romani (descritti da pallini colorati in blu).

2. Pom-78, Pom-110 e Pom-112 ricadono al di fuori della variabilità genetica prettamente Imperiale, localizzandosi in prossimità delle popolazioni del BA e dell'IA Italiano (*Italy_BA* e *Italy_IA*). Questa vicinanza è particolarmente marcata per Pom-110. Pom-78 si posiziona invece vicino alle popolazioni delle province Romane Inglesi (*England_Roman*) e agli individui Imperiali francesi e portoghesi (*France_Imperial* e *Portugal_Imperial*). Infine, Pom-112 sembra essere al limite della variabilità Imperiale italiana, non troppo distante dalle popolazioni Imperiali francesi.
3. Una differenza rispetto alla PCA con il dato pseudo-aploide si ha nella localizzazione di Pom-41 che risultava al di fuori del cluster Imperiale, in prossimità dell'attuale variabilità del nord Africa insieme agli individui delle province Romane Spagnole (*Spain_Roman*), in quest'analisi assenti poiché dati di cattura. L'attuale posizione di Pom-41 (sulla variabilità Imperiale) può quindi essere dovuta a un maggior recupero di posizioni usate per l'analisi grazie all'approccio *whole-genome*.
4. Vi sono infine degli individui (Pom-77, Pom-104, Pom-107, Pom-124) che tendono a rimanere al di fuori della variabilità Imperiale, sebbene non risultino vicini alle altre popolazioni antiche usate per i confronti.

4.2.5.2 NGSAdmix

Per analizzare con maggiore dettaglio la diversità genomica osservata in PCA, è stata effettuata un'analisi di *Admixture* condotta su tutti gli individui antichi presenti nel dataset (la cui selezione è stata effettuata seguendo i criteri descritti nel *paragrafo 3.4.8.1*) includendo, a differenza della prima anche le popolazioni ancestrali. L'analisi ha sfruttato le GL calcolate su 1246484 posizioni autosomiche polimorfiche. Sebbene il numero di popolazioni sorgente fornite a NgsAdmix vari da 2 a 10 (-K2 ÷ -K10), è stato messo in grafico l'*output* generato assumendo che la popolazione di Pompei derivi da 5 gruppi ancestrali (K5; *Figura 16*) poiché permette di individuare i *pattern* di variabilità tra individui e popolazioni con maggiore chiarezza.

In *Figura 16* ogni rettangolo rappresenta un individuo appartenente alla popolazione descritta dal *label*. Ciascun rettangolo è caratterizzato da uno o più colori, in proporzioni differenti, rappresentanti la componente genomica o le componenti massimizzate nel proprio DNA.

L'analisi di *Admixture* ha permesso di confermare i *patterns* di variabilità già individuati dall'analisi delle componenti principali (PCA).

Sebbene la componente Neolitica dei primi agricoltori Anatolici (*Anatolia_N*) sia descritta dal colore giallo, nel CA Italiano (*Italy_CA*, periodo in cui nel genoma degli individui sono presenti le sole

componenti riferibili al Neolitico e ai *WHG*, qui in azzurro) l'*ancestry* Neolitica è accompagnata da un'ulteriore componente (verde) che in quest'analisi non è massimizzata in nessuno degli individui più antichi.

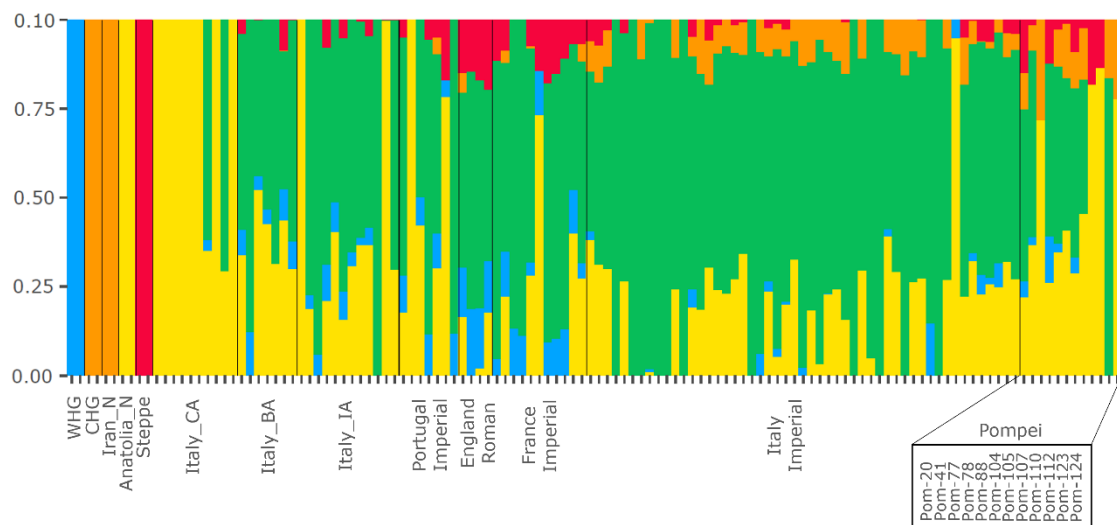


Figura 16 Grafico dei risultati ottenuti dell'analisi di *Admixture* con K5 (5 popolazioni sorgente). A sinistra del grafico sono presenti le popolazioni ancestrali, seguono le popolazioni antiche usate nei confronti in ordine temporale e infine i 12 individui *deep shotgun* di Pompei.

Il passaggio dal CA al BA è segnato dalla comparsa, nel genoma di alcuni individui, della componente riferibile alla steppa (*Steppe*, indicata in rosso). Come già osservato in PCA (*Figura 15*), anche l'analisi in *Figura 16* mostra il mantenimento nel corso dell'IA della stessa struttura genetica del periodo precedente. Il periodo Imperiale, al di fuori del territorio Italiano, si caratterizza per la forte presenza della componente della steppa, più in Francia e nelle colonie Romane Inglesi (*France Imperial*, *England Roman*) che in Portogallo (*Portugal Imperial*). Questo *pattern* era già visibile in PCA (*Figura 15*) in quanto tutte le popolazioni su menzionate si dispongono le une vicino alle altre sulla destra del grafico. Un evidente cambio nella struttura genomica della popolazione del territorio Italiano si ha nel corso del periodo Imperiale. In *Figura 16*, gli individui riferibili all'*Italy Imperial* sono caratterizzati da una scarsa presenza della componente della steppa e dalla comparsa di una nuova componente genomica riconducibile alle popolazioni Neolitiche Iraniane (*Iran_N*, qui in arancione). Sebbene essa si trovi, in misura minore, anche in qualche individuo estero aumenta però significativamente tra gli individui Italiani. Questa diversità è accompagnata da uno *shift* in PCA (*Figura 15*) verso la sinistra del grafico. Ciò che si osserva a Pompei dall'analisi di *Admixture* è un insieme di tutto quello che è stato trovato non solo nello stesso periodo al di fuori della penisola, ma anche in periodi diversi del medesimo territorio.

Pom-41, Pom-88 e Pom-123 possiedono un profilo in linea con la variabilità genetica già individuata tra gli *Italy_Imperial*. Anche Pom-104 e Pom-107 sembrano avere un'ancestry tipicamente Imperiale anche se, a differenza dei primi, la loro posizione in PCA risulta distante dal *cluster* Imperiale principale (in loro risulta assente la componente dei WHGs). Gli individui Pom-105 e Pom-20, sebbene abbiano un profilo genetico in linea con la variabilità italiana Imperiale, sono caratterizzati da una componente della steppa pari o predominante a quella Neolitica Iraniana; infatti, in PCA nonostante non siano lontani dalla variabilità di quel periodo tendono ad avvicinarsi a qualche individuo *France_Imperial* in cui è attestata una maggiore presenza dell'ancestry della steppa. Peculiare il profilo di Pom-78 in cui risulta assente la componente *Iran_N* e la cui struttura genetica coincide con quella degli individui delle province Romane Inglesi (*England_Roman*) e del periodo Imperiale francese (*France_Imperial*), come già suggerito dalla PCA in *Figura 15*. Nel genoma degli individui Pom-77 e Pom-124 viene massimizzata la componente *Anatolia_N* e *Iran_N*, situazione completamente opposta a quella di Pom-110 e Pom-112 in cui la componente arancione è rimpiazzata completamente da quella rossa delle *Steppe*. Lo stato appena descritto trova infatti conferma nella posizione di quest'ultimi (Pom-110 e Pom-112) in PCA, essi risultano al di fuori della variabilità Imperiale e il loro profilo ricorda quello delle popolazioni del BA e dell'IA Italiano. Interessante notare come in quest'analisi, quattro degli individui outlier in *Figura 15* (Pom-77, Pom-104, Pom-107, Pom-124) sebbene siano tutti disposti alla sinistra del grafico mostrano profili genomici non omogenei fra loro. Punto in comune fra essi sembra però essere la totale, o quasi, assenza della componente della steppa, ricorrente invece negli individui spostati verso la destra della PCA in *Figura 15*. In futuro, sarà comunque necessario ripetere l'analisi ampliando il numero di popolazioni ancestrali per meglio evidenziare le differenze nel genoma degli individui studiati.

I risultati mostrati e ottenuti a partire dal calcolo delle GL dimostrano che esse sono quindi un'alternativa valida e statisticamente più robusta per poter effettuare studi genomici di popolazioni antiche sequenziate a basso *coverage*.

4.3 La Stanza degli Scheletri

Precedenti analisi, eseguite al Dipartimento di Biologia dell'Università degli Studi di Firenze, con approccio forense (tipizzazione dei loci Short Tandem Repeat) hanno dimostrato, a partire da 12 resti umani, la presenza di 10 individui all'interno della "Stanza degli Scheletri" nel complesso "Casa del Giardino". Infatti, i resti umani di Pom-4 e Pom-9 risultano appartenere alla stessa persona così come appartengono allo stesso individuo i resti di Pom-5 e Pom-6.

Le analisi antropologiche hanno identificato la presenza, all'interno della stanza, di 4 adulti e 6 subadulti. Interessante notare, come già riportato nel *paragrafo 3.2*, la posizione in cui i resti umani sono stati rinvenuti nonostante il sito abbia subito delle manomissioni nel corso dei secoli: gli adulti in prossimità della porta, mentre i subadulti sul fondo della stanza.

Successivamente al primo sequenziamento esplorativo, solo i 6 individui con percentuale di DNA endogeno superiore al 10% sono stati selezionati per un secondo sequenziamento al fine di raggiungere un *coverage* medio di 0.1x. Dato il particolare contesto di scavo, essi sono stati trattati diversamente dai 12 individui *deep shotgun* (in precedenza analizzati) poiché l'obiettivo dello studio non era tanto l'inferenza demografica, quanto quello di verificare eventuali relazioni di parentela tra gli adulti e i subadulti.

Il trattamento dei dati grezzi con la *pipeline* descritta nel *paragrafo 3.4.2* ha portato alla ricostruzione dei genomi le cui principali statistiche sono riportate in *Tabella 16*.

ID campione	Reads grezze	Reads mappanti	DNA endogeno (%)	Coverage medio	Coverage 1X (%)
Pom-1	100365528	7341543	27.768	0.1185	10.73
Pom-2	140110200	4839696	13.072	0.0763	7.17
Pom-3	62713832	6712805	46.02	0.1093	10.11
Pom-4/9	44097032	6284752	62.912	0.096	8.95
Pom-5/6	61356854	7849451	57.732	0.1191	10.91
Pom-12	77471484	6707855	35.452	0.1087	10.02

Tabella 16 Principali statistiche per i 6 campioni processati come descritto nel *paragrafo 3.4.2*

Sempre utilizzando il bam file ricalibrato è stato determinato il sesso come descritto nel *paragrafo 3.4.3* che ha restituito i risultati mostrati in *Tabella 17*. A tutti gli adulti (Pom-2, Pom-3 e Pom-12) è stato attribuito il sesso femminile, mentre i subadulti (Pom-1, Pom-4 e Pom-5) risultano sia femmine che maschi.

ID campioni	NchrY+NchrX	NchrY	Ry	SE	95% IC	Assegnazione
Pom-1	189092	14598	0.0772	0.0006	0.076-0.0784	XY
Pom-2	227538	169	0.0007	0.0001	0.0006-0.0009	XX
Pom-3	325949	161	0.0005	0	0.0004-0.0006	XX
Pom-4/9	170448	13054	0.0766	0.0006	0.0753-0.0778	XY
Pom-5/6	379544	207	0.0005	0	0.0005-0.0006	XX
Pom-12	321872	193	0.0006	0	0.0005-0.0007	XX

Tabella 17 Risultati relativi al sesso determinato su base molecolare.

Per i due individui di sesso maschile è stata dimostrata l'assenza di contaminazione nucleare ($-0.002057 \pm 2.09E-03$ e $-0.002494 \pm 2.54E-03$) usando il metodo descritto nel *paragrafo 3.4.4*. Inoltre, è stato identificato l'aplogruppo del cromosoma Y come descritto nel *paragrafo 3.4.5*, con la modifica del parametro $-r$ 2 che non aveva inizialmente prodotto nessun risultato a causa del basso *coverage* di sequenziamento. Impostando la *flag* a 1 ($-r$ 1) sono stati determinati gli aplogruppi mostrati in *tab Tabella 18*.

ID campioni	Aplogruppo chrY	Marcatoti validi
Pom-1	R1b1a1b	4114
Pom-4/9	E1b1b1b~	3598

Tabella 18 Aplogruppi del cromosoma Y identificati abbassando il numero minimo di *reads* mappanti sui marcatori chrY a 1.

Le analisi sul genoma mitocondriale sono state eseguite a seguito degli esperimenti di cattura realizzati in precedenza al Dipartimento di Biologia dell'Università degli Studi di Firenze tramite l'impiego di sonde specifiche per il DNA mitocondriale (Maricic et al., 2010). È stato impiegato il *software* ContamMix (Fu et al., 2013; Fu et al., 2014) per stimare la contaminazione sul genoma catturato ed è stato determinato l'aplogruppo utilizzando il *software* Haplofind (Vianello et al., 2013) che è stato poi confermato mediante l'impiego di Mitomaster (Brandon et al., 2009); i risultati dell'analisi mitocondriale sono riportati in *Tabella 19*.

ID campione	Reads on MT	Coverage MT	Contaminazione mtDNA	Aplogruppo mtDNA
Pom-1	240080	809.9197	99.36	T2b7a2
Pom-2	129871	433.6295	99.61	H4a1a1a
Pom-3	131635	446.4671	97.97	K1a2
Pom-4/9	202959	632.0716	98.91	U7a
Pom-5/6	229167	708.9081	99.21	I2
Pom-12	243753	822.3547	99.16	I2

Tabella 19 Statistiche sul genoma mitocondriale a seguito di precedenti lavori eseguiti all'Università degli Studi di Firenze. Sono riportate le *reads* finali mappanti sulla rCRS, il *coverage* medio, l'autenticità del genoma endogeno (95% IC) e l'aplogruppo.

I 6 genomi *shotgun* 0.1x (Tabella 16) sono stati tipizzati per gli *SNPs-array* HO e 1240K come descritto nel paragrafo 3.4.6. In Tabella 20 sono riportate le coperture degli individui per i due pannelli di *SNPs* utilizzati.

ID campioni	SNPs coperti 1240K	SNPs coperti HO
Pom-1	175159	95680
Pom-2	115731	62733
Pom-3	160598	86089
Pom-4/9	146049	78261
Pom-5/6	175670	94027
Pom-12	159543	85803

Tabella 20 *SNPs* coperti in seguito alla genotipizzazione effettuata sia sul pannello *Human Origins* (HO 597573 *SNPs* totali) che sul 1240K (1233013 *SNPs* totali).

L'analisi di parentela sui genomi *shotgun*, utilizzando il software READ (Monroy Kuhn et al., 2018) studiato per genomi antichi a basso *coverage*, ha permesso l'identificazione di una relazione di primo grado tra Pom-5/6 e Pom-12 (Figura 17), una donna e una bambina che, condividendo lo stesso aplogruppo mitocondriale (Tabella 19), potrebbero essere madre-figlia.

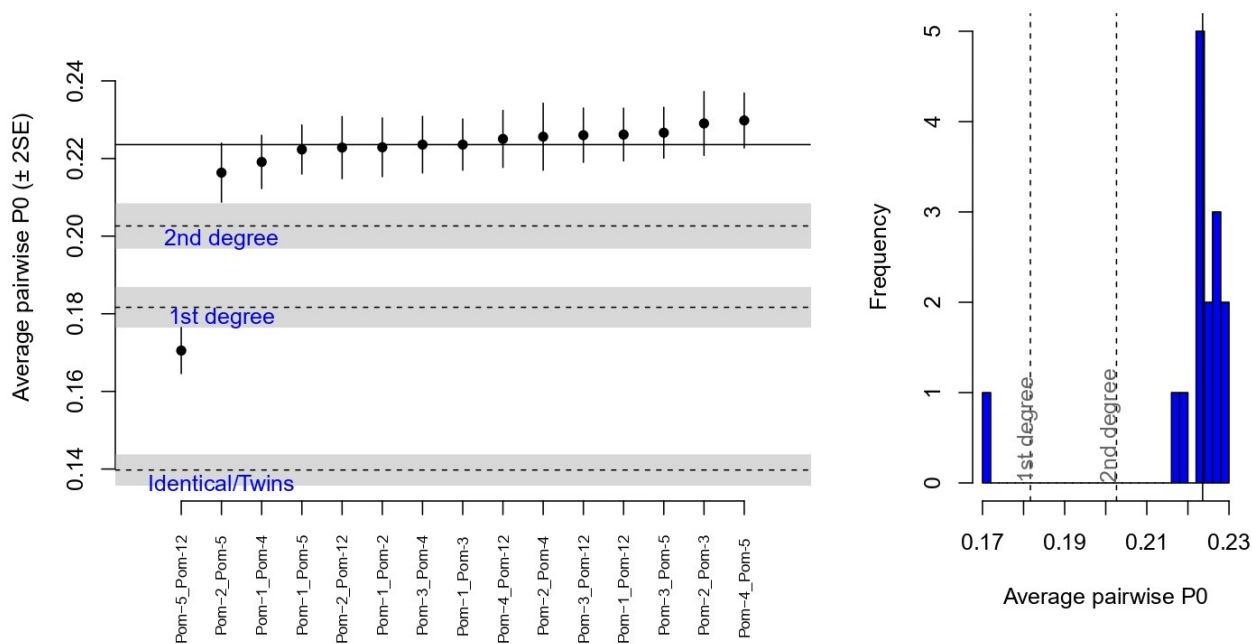


Figura 17 Analisi di parentela effettuata come descritto nel *paragrafo 3.4.7*. Nel grafico di sinistra è mostrato il tipo di parentela tra le coppie di individui, in quello di destra la proporzione di alleli differenti (P0).

La PCA è stata costruita utilizzando le attuali popolazioni dell’Eurasia occidentale e del Nord Africa sulle quali sono state proiettate le popolazioni antiche di confronto e tutti e 6 gli individui della “Stanza degli Scheletri” come già descritto nel *paragrafo 3.4.6*.

È interessante notare in *Figura 18* come individui ritrovati nel medesimo luogo siano disposti in PCA su tutti i clini di variabilità delle attuali popolazioni Eurasiatiche, a indicazione di una grande variabilità genetica all’interno di un gruppo molto ristretto di individui a Pompei. Inoltre, va evidenziato come gli individui legati da una relazione di primo grado, quale quella madre-figlia (Pom-12 e Pom-5/6), siano posizionati in punti distanti della PCA suggerendo un’*ancestry* paterna completamente differente da quella materna. Pom-12 si posiziona sulla variabilità delle attuali popolazioni dell’Europa dell’ovest e nel confronto con le popolazioni antiche risulta lontana dalla variabilità delle popolazioni imperiali del territorio Italiano avvicinandosi invece ai coevi individui francesi e delle colonie inglesi. Pom-5, invece, si colloca sulla variabilità delle attuali popolazioni dell’Italia centrale a breve distanza dagli altri imperiali italiani.

In futuro, al fine di completare la ricerca di parentele tra gli individui all’interno della stanza, si prevede la ricostruzione del genoma dei campioni mancanti (Pom-7, Pom-8, Pom-10 e Pom-11, tutti a bassa

preservazione del DNA endogeno) applicando il metodo di cattura degli SNPs del pannello 1240K, in quanto risulta sufficiente una percentuale di endogeno superiore allo 0.15% per operare l'arricchimento. Per alcuni di essi è inoltre previsto un secondo campionamento da nuovi resti umani.

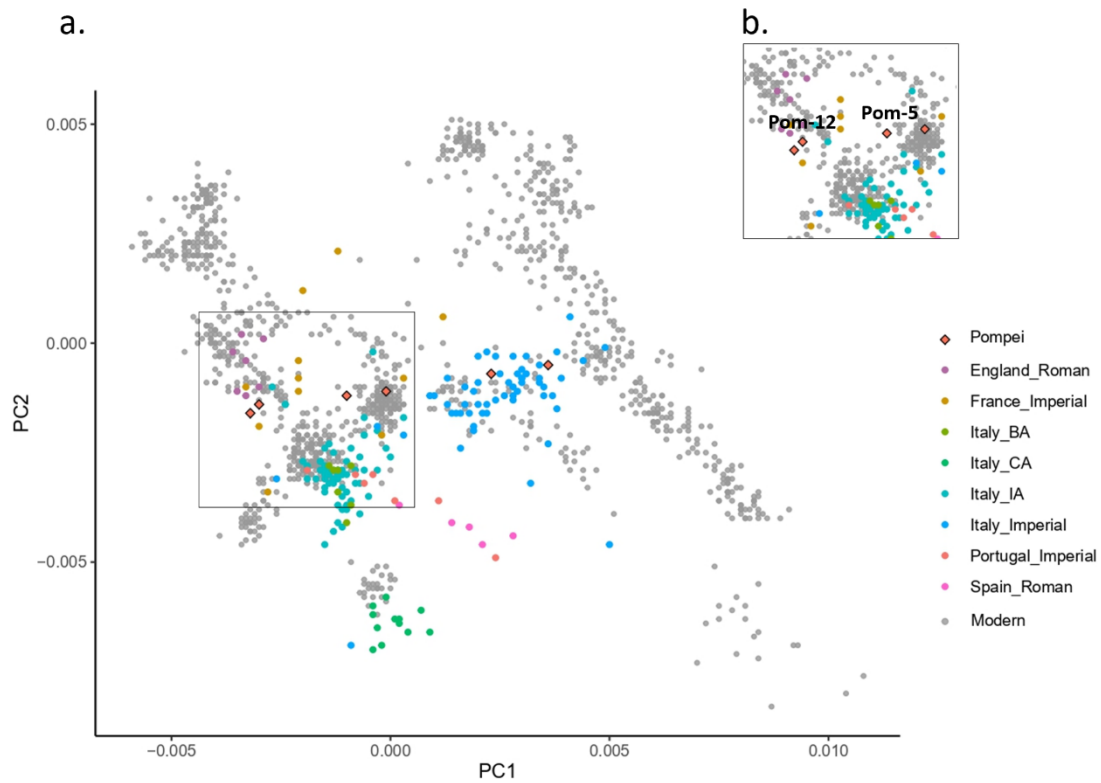


Figura 18 a. Analisi delle Componenti Principali (PCA) realizzata a partire dalla variabilità genomica di 61 popolazioni dell'Eurasia occidentale e dell'Africa del nord genotipizzate per il pannello HO in cui sono stati proiettati i 6 individui della "Stanza degli Scheletri" e i campioni antichi di confronto. **b.** Dettaglio della PCA in cui sono evidenziate madre (Pom-12) - figlia (Pom-5).

5 CONCLUSIONI

Il presente progetto di ricerca ha visto il campionamento di più di 170 resti umani provenienti da diverse aree del Parco Archeologico di Pompei e riferibili alle vittime dell'eruzione del complesso Somma-Vesuvio avvenuta nel 79 e.c. Il campionamento dei reperti è stato realizzato con l'obiettivo di recuperare e studiare l'aDNA al fine di chiarire, nel caso della "Stanza degli Scheletri", la presenza di relazioni di parentela sospettate dagli antropologi successivamente al ritrovamento dei resti umani e, più in generale, indagare la variabilità genetica degli individui che, in quel particolare momento storico, vivevano o si trovavano nell'antica città di Pompei.

I primi risultati, ottenuti dal sequenziamento esplorativo di un numero così elevato di resti umani appartenenti alle vittime dell'eruzione, confermano la presenza di DNA endogeno altamente conservato e preservato nel corso degli anni.

Tramite un approccio multidisciplinare, applicato ai resti umani della "Stanza degli Scheletri", siamo riusciti a dare risposta a molte domande avanzate in fase di scoperta del sito. Solo donne e bambini riuscirono a trovare un provvisorio riparo all'interno della stanza della *domus* nelle ultime fasi dell'eruzione. Nonostante la profondità di sequenziamento per i 6 individui fosse inferiore all'1x, sono stati applicati con successo metodi e strumenti ideati per l'analisi di genomi degradati a basso *coverage*, mettendo in luce che due di essi risultano legati da una parentela di primo grado, anche se ad oggi non si può escludere la presenza di ulteriori legami di sangue tra altri individui data l'impossibilità di analizzare la totalità del campione. Infine, è stata individuata grande variabilità genetica in un numero così ristretto di persone tanto che "madre-figlia" risultano localizzate in punti diversi della PCA, la prima sovrapposta alle attuali popolazioni francesi e inglesi non lontana dagli individui delle province Romane Inglesi, la seconda in prossimità della variabilità individuata in Italia nel periodo Imperiale.

Gli approcci allo studio dei genomi sono stati di due tipi e hanno dato esiti molto simili, ma quello basato sullo studio dell'intero genoma e sul calcolo delle *genotype likelihood* (GL) ha fornito un maggiore supporto statistico alle conclusioni.

Poiché la maggior parte dei genomi antichi, nell'era delle NGS, viene ricostruita con una profondità di sequenziamento medio-bassa (a causa della scarsa presenza del DNA endogeno o per la bassa disponibilità economica nel sequenziamento di un numero elevato di campioni), il calcolo delle *likelihood* rappresenta un buon metodo per la loro analisi.

Tramite le *likelihood* è stato possibile estendere l'analisi a un numero maggiore di posizioni genomiche, in parte escluse dai pannelli di riferimento più utilizzati negli studi di antropologia molecolare, caratterizzando più nel dettaglio gli individui in esame e arrivando quindi a conclusioni più robuste. Inoltre, sono stati considerati fattori che, diversamente, avrebbero introdotto delle incertezze nella

chiamata dei genotipi (bassa qualità di lettura dei nucleotidi, bassa copertura della regione con rischio di campionamento di un solo cromosoma in organismi diploidi, bassa qualità del mappaggio ecc.).

In questo lavoro, per la prima volta in modo concreto, viene dimostrata la possibilità di applicare metodi di analisi basati sulle GL per realizzare studi di variabilità e analisi dell'*ancestry* in popolazioni antiche di forte interesse storico, i cui genomi sono necessariamente ricostruiti a basso *coverage*.

Le analisi popolazionistiche dimostrano, su un campione così esteso della popolazione, che la variabilità genetica intra-popolazionistica nell'antica città Pompei, in confronto con quella già documentata nello stesso periodo nella capitale dell'Impero (e nei dintorni di essa) risulta elevata, sebbene Pompei fosse una città molto più piccola di Roma e abbia visto la sua crescita economica e culturale in seguito al trasferimento della nobiltà della capitale. Gli individui a Pompei non solo sono differenti gli uni dagli altri, come dimostrato dall'ampia distribuzione dei campioni in PCA e dall'analisi di *Admixture*, ma il loro profilo genetico risulta differente da quello degli altri individui Imperiali della penisola, quest'ultimi geneticamente molto più omogenei. Infatti, a differenza degli altri campioni di epoca imperiale, che non mostrano vicinanza genetica con le popolazioni delle province straniere dell'Impero (se non in rari casi), alcuni degli individui di Pompei sembrano condividere con esse il loro profilo genico, in particolar modo con le popolazioni inglesi e francesi. Infine, in alcuni casi è stato osservato a Pompei il mantenimento del profilo genomico che caratterizzava le popolazioni Italiche preromane.

Questo dimostra la dinamicità di Pompei, importante città di porto nel Mediterraneo, in cui la presenza di individui con origini diverse e non comuni, considerate le popolazioni dei territori vicini, può essere il frutto dei contatti, antichi e duraturi o prossimi al periodo dell'eruzione, delle popolazioni romane con popolazioni interne e/o esterne all'Impero per fini economici, lavorativi e militari.

A riguardo, sarà interessante in futuro esaminare i risultati relativi all'analisi degli isotopi, i quali forniscono indicazioni riguardo al luogo di nascita e crescita degli individui. In questo modo, sarà possibile capire in che misura la diversità della popolazione pompeiana sia il frutto di migrazioni avvenute nell'ultima generazione prima dell'eruzione (documentate dallo studio degli isotopi) oppure precedenti (che hanno lasciato traccia nel genoma degli individui).

Bibliografia

- Abulafia D, 2011. *The Great Sea: A Human History of the Mediterranean*. Oxford University Press
- Ammerman AJ, Cavalli-Sforza LL, *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press, 1984. JSTOR, <http://www.jstor.org/stable/j.ctt7zvqz7>
- Antonio ML, et al., Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*. 2019 Nov 8;366(6466):708-714. doi: 10.1126/science.aay6826. PMID: 31699931; PMCID: PMC7093155
- Antonio ML, et al., Stable population structure in Europe since the Iron Age, despite high mobility. *bioRxiv*, 2023: p. 2022.05.15.491973. doi: <https://doi.org/10.1101/2022.05.15.491973>
- Allentoft ME, et al., Population genomics of Bronze Age Eurasia. *Nature*. 2015 Jun 11;522(7555):167-72. doi: 10.1038/nature14507. PMID: 26062507
- Avila-Arcos M, et al., Comparative Performance of Two Whole Genome Capture Methodologies on Ancient DNA Illumina Libraries. *Methods in Ecology and Evolution*. 2015. 6. 10.1111/2041-210X.12353
- Barlow A, et al., Partial genomic survival of cave bears in living brown bears. *Nat Ecol Evol*. 2018 Oct;2(10):1563-1570. doi: 10.1038/s41559-018-0654-8. Epub 2018 Aug 27. PMID: 30150744; PMCID: PMC6590514
- Brace S, et al., Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat Ecol Evol*. 2019 May;3(5):765-771. doi: 10.1038/s41559-019-0871-9. Epub 2019 Apr 15. Erratum in: *Nat Ecol Evol*. 2019 May 8;: PMID: 30988490; PMCID: PMC6520225
- Brandon MC, et al., MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum Mutat*. 2009 Jan;30(1):1-6. doi: 10.1002/humu.20801. PMID: 18566966; PMCID: PMC4292933
- Briggs AW, et al., Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*. 2007 Sep 11;104(37):14616-21. doi: 10.1073/pnas.0704665104. Epub 2007 Aug 21. PMID: 17715061; PMCID: PMC1976210
- Broodbank C, 2013. *The making of the Middle Sea: a history of the Mediterranean from the beginning to the emergence of the classical world*. Oxford; New York: Oxford University Press
- Brotherton P, et al., Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res*. 2007;35(17):5717-28. doi: 10.1093/nar/gkm588. Epub 2007 Aug 22. PMID: 17715147; PMCID: PMC2034480
- Carpenter ML, et al., Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American journal of human genetics*. 2013. 93(5):852-64. PUBMED: 24568772 DOI:S0002-9297(13)00459-X
- Cipollaro M, et al., Ancient DNA in human bone remains from Pompeii archaeological site. *Biochem Biophys Res Commun*. 1998 Jun 29;247(3):901-4. doi: 10.1006/bbrc.1998.8881. PMID: 9647790
- Cooper A, Poinar HN. Ancient DNA: do it right or not at all. *Science*. 2000 Aug 18;289(5482):1139. doi: 10.1126/science.289.5482.1139b. PMID: 10970224

Dabney J, Meyer M, Pääbo S. Ancient DNA damage. *Cold Spring Harb Perspect Biol.* 2013 Jul 1;5(7):a012567. doi: 10.1101/cshperspect.a012567. PMID: 23729639; PMCID: PMC3685887.

Dabney J, et al., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A.* 2013 Sep 24;110(39):15758-63. doi: 10.1073/pnas.1314445110. Epub 2013 Sep 9. PMID: 24019490; PMCID: PMC3785785

Danecek P, et al., Twelve years of SAMtools and BCFtools. *GigaScience.* Volume 10. Issue 2. February 2021. giab008. <https://doi.org/10.1093/gigascience/giab008>

de Barros Damgaard P, et al., The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science (New York, N.Y.)*, 2018. 360(6396): p. eaar7711

De Franciscis A, La casa di Caius Iulius Polibius. *Rivista di studi Pompeiani.* 1988. II, 15–36

Di Bernardo G, et al., Ancient DNA and family relationships in a Pompeian house. *Ann Hum Genet.* 2009 Jul;73(Pt 4):429-37. doi: 10.1111/j.1469-1809.2009.00520.x. PMID: 19523150

Fernandes DM, et al., The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat Ecol Evol.* 2020 Mar;4(3):334-345. doi: 10.1038/s41559-020-1102-0. Epub 2020 Feb 24. Erratum in: *Nat Ecol Evol.* 2020 May;4(5):764. PMID: 32094539; PMCID: PMC7080320

Fort J, Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc Natl Acad Sci U S A.* 2012 Nov 13;109(46):18669-73. doi: 10.1073/pnas.1200662109. Epub 2012 Oct 29. PMID: 23112147; PMCID: PMC3503213

Fort J, Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. *J R Soc Interface.* 2015. 12:1–8.

Fu Q, et al., DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A.* 2013 Feb 5;110(6):2223-7. doi: 10.1073/pnas.1221359110. Epub 2013 Jan 22. PMID: 23341637; PMCID: PMC3568306

Fu Q, et al., Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature.* 2014 Oct 23;514(7523):445-9. doi: 10.1038/nature13810. PMID: 25341783; PMCID: PMC4753769

Furtwängler A, Reiter E, Neumann GU et al., Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Sci Rep* 8, 14075 (2018). <https://doi.org/10.1038/s41598-018-32083-0>

Gamba C, Jones E, Teasdale M, et al., Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5, 5257 (2014). <https://doi.org/10.1038/ncomms6257>

Gilbert MT, Bandelt HJ, Hofreiter M, Barnes I. Assessing ancient DNA studies. *Trends Ecol Evol.* 2005 Oct;20(10):541-4. doi: 10.1016/j.tree.2005.07.005. Epub 2005 Jul 21. PMID: 16701432

Gilbert MT, et al., DNA from pre-Clovis human coprolites in Oregon, North America. *Science.* 2008 May 9;320(5877):786-9. doi: 10.1126/science.1154116. Epub 2008 Apr 3. PMID: 18388261

González-Fortes G, et al., Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. *Current Biology.* June 19 2017. Volume 27, issue 12, P1801-1810.E10

Green RE, et al., A draft sequence of the Neandertal genome. *Science.* 2010 May 7;328(5979):710-722. doi: 10.1126/science.1188021. PMID: 20448178; PMCID: PMC5100745

Günther T, et al., Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences of the United States of America*. 2015. 112. 10.1073/pnas.1509851112

Günther T, Jakobsson M, Genes mirror migrations and cultures in prehistoric Europe—a population genomic perspective. *Curr Opin Genet Dev*. 2016 Dec; 41:115-123. doi: 10.1016/j.gde.2016.09.004. Epub 2016 Sep 27. PMID: 27685850

Haak W, et al., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 2015 522, 207–211

Harris EE, Demic and cultural diffusion in prehistoric Europe in the age of ancient genomes. *Evol Anthropol*. 2017 Sep;26(5):228-241. doi: 10.1002/evan.21545. PMID: 29027332

Higuchi R, et al., DNA sequences from the quagga, an extinct member of the horse family. *Nature*. 1984 Nov 15-21;312(5991):282-4. doi: 10.1038/312282a0. PMID: 6504142

Hofreiter M, et al., Ancient DNA. *Nat Rev Genet*. 2001 May;2(5):353-9. doi: 10.1038/35072071. PMID: 11331901.

Hofreiter M, et al., The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays*. 2015 Mar;37(3):284-93. doi: 10.1002/bies.201400160. Epub 2014 Nov 21. PMID: 25413709

Jensen TZT, et al., A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat Commun*. 2019. 10, 5520. <https://doi.org/10.1038/s41467-019-13549-9>

Jones ER, et al., Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun*. 2015 Nov 16; 6:8912. doi: 10.1038/ncomms9912. PMID: 26567969; PMCID: PMC4660371

Jones ER, et al., The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology*. Volume 27, Issue 4. 20 February 2017. Pages 576-582

Jónsson H, et al., mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 2013 Jul 1; 29(13):1682-4. doi: 10.1093/bioinformatics/btt193. Epub 2013 Apr 23. PMID: 23613487; PMCID: PMC3694634

Jun G, et al., An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. 2015 April. *Genome research* gr-176552.114

Kılınc GM, et al., The Demographic Development of the First Farmers in Anatolia. *Curr Biol*. 2016 Oct 10;26(19):2659-2666. doi: 10.1016/j.cub.2016.07.057. Epub 2016 Aug 4. PMID: 27498567; PMCID: PMC5069350

Knapp M, Hofreiter M, Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. *Genes (Basel)*. 2010. Jul 28; 1(2):227-43. doi: 10.3390/genes1020227. PMID: 24710043; PMCID: PMC3954087

Korneliussen TS, Albrechtsen A, Nielsen R, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 2014. 15, 356. <https://doi.org/10.1186/s12859-014-0356-4>

Krause J, et al., A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol*. 2010 Feb 9; 20(3):231-6. doi: 10.1016/j.cub.2009.11.068. Epub 2009 Dec 31. PMID: 20045327

Lazaridis I, et al., Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 2014. 513(7518): p. 409-413

Lazaridis I, et al., Genomic insights into the origin of farming in the ancient Near East. *Nature*. 2016 Aug 25; 536(7617):419-24. doi: 10.1038/nature19310. Epub 2016 Jul 25. PMID: 27459054; PMCID: PMC5003663

Li H, et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002

Li H, Durbin R, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. Epub 2009 May 18. PMID: 19451168; PMCID: PMC2705234

Marchi N, et al., The genomic origins of the world's first farmers. *Cell*. 2022. Vol 185, Issue 11. 1842-1859.e18. <https://doi.org/10.1016/j.cell.2022.04.008>

Maricic T, Whitten M, Pääbo S, Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 2010 Nov 16;5(11):e14004. doi: 10.1371/journal.pone.0014004. PMID: 21103372; PMCID: PMC2982832

Mathieson I, et al., Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*. 2015 Dec 24;528(7583):499-503. doi: 10.1038/nature16152. Epub 2015 Nov 23. PMID: 26595274; PMCID: PMC4918750

Matsuo S, et al., Degradation of DNA in dried tissues by atmospheric oxygen. *Biochem. Biophys. Res. Commun.* 1995. 208, 1021–1027

McKenna A, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010 Sep; 20(9):1297-303. doi: 10.1101/gr.107524.110. Epub 2010 Jul 19. PMID: 20644199; PMCID: PMC2928508

Meisner J, Albrechtsen A, Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* Volume 210. Issue 2, 1 October 2018. Pages 719–731, <https://doi.org/10.1534/genetics.118.301336>

Meyer M, Kircher M, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2010 Jun;2010(6):pdb.prot5448. doi: 10.1101/pdb.prot5448. PMID: 20516186

Meyer M, et al., A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 2012 Oct 12;338(6104):222-6. doi: 10.1126/science.1224344. Epub 2012 Aug 30. PMID: 22936568; PMCID: PMC3617501

Modi A, Vai S, et al., More data on ancient human mitogenome variability in Italy: new mitochondrial genome sequences from three Upper Palaeolithic burials. *ANNALS OF HUMAN BIOLOGY*, 2021, VOL. 48, NO. 3, 213–222

Monroy Kuhn JM, Jakobsson M, Günther T, Estimating genetic kin relationships in prehistoric populations. *PLoS One*. 2018 Apr 23;13(4):e0195491. doi: 10.1371/journal.pone.0195491. PMID: 29684051; PMCID: PMC5912749

Moots HM, Antonio M, Sawyer S, et al. A genetic history of continuity and mobility in the Iron Age central Mediterranean. *Nat Ecol Evol* 2023. <https://doi.org/10.1038/s41559-023-02143-4>

Mullis K, et al., Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51 Pt 1:263-73. doi: 10.1101/sqb.1986.051.01.032. PMID: 3472723

Narasimhan VM, et al., The formation of human populations in South and Central Asia. *Science.* 2019 Sep 6;365(6457):eaat7487. doi: 10.1126/science.aat7487. PMID: 31488661; PMCID: PMC6822619

Olalde I, et al., The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature,* 2018. 555(7695): p. 190-196

Olalde I, et al., The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363,1230-1234. 2019. DOI:10.1126/science.aav4040

Omrak A, et al., Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr Biol* 2016, 26:270-275

Özdoğan M, Archaeological evidence on the westward expansion of farming communities from eastern anatolia to the aegean and the balkans. *Curr Anthropol* 2011, 52:S415-S430

Pääbo S. Molecular cloning of Ancient Egyptian mummy DNA. *Nature.* 1985. Apr 18-24;314(6012):644-5. doi: 10.1038/314644a0. PMID: 3990798

Pääbo S, Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A.* 1989. 86(6): p. 1939-43

Palkopoulou E, et al., Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol.* 2015. May 18; 25(10):1395-400. doi: 10.1016/j.cub.2015.04.007. Epub 2015 Apr 23. PMID: 25913407; PMCID: PMC4439331

Patterson N, Price AL, Reich D, Population Structure and Eigenanalysis. *PLOS Genetics.* 2006. 2(12): e190. <https://doi.org/10.1371/journal.pgen.0020190>

Peltzer A, et al., EAGER: efficient ancient genome reconstruction. *Genome Biol.* 2016 Mar 31; 17:60. doi: 10.1186/s13059-016-0918-z. PMID: 27036623; PMCID: PMC4815194

Pilli E, et al., Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. *PLoS One.* 2013;8(1):e52524. doi: 10.1371/journal.pone.0052524. Epub 2013 Jan 25. PMID: 23372650; PMCID: PMC3556025

Pinhasi R, et al., Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One.* 2015 Jun 18;10(6):e0129102. doi: 10.1371/journal.pone.0129102. PMID: 26086078; PMCID: PMC4472748.

Plinio, *episola VI 16, VI 20*

Posth C, et al., Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Current Biology.* Volume 26. Issue 6, 21 March. 2016. Pages 827-833, ISSN 0960-9822

Posth C, et al., The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Sci. Adv.* 2021. 7, eabi 7673. DOI: 10.1126/sciadv.abi7673

Prüfer K, et al., The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014 Jan 2;505(7481):43-9. doi: 10.1038/nature12886. Epub 2013 Dec 18. PMID: 24352235; PMCID: PMC4031459

Purcell S, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep;81(3):559-75. doi: 10.1086/519795. Epub 2007 Jul 25. PMID: 17701901; PMCID: PMC1950838

Quinlan AR, Hall IM, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010. 26, 6, pp. 841–842

Raghavan M, et., Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* 2014. Jan 2; 505(7481):87-91. doi: 10.1038/nature12736. Epub 2013 Nov 20. PMID: 24256729; PMCID: PMC4105016

Ralf A, et al., Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Molecular Biology and Evolution.* 2018. 35(5): p. 1291-1294

Rasmussen M, et al., An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science.* 2011 Oct 7; 334(6052):94-8. doi: 10.1126/science.1211177. Epub 2011 Sep 22. PMID: 21940856; PMCID: PMC3991479

Reich D, et al., Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature.* 2010 Dec 23;468(7327):1053-60. doi: 10.1038/nature09710. PMID: 21179161; PMCID: PMC4306417

Renaud G, et al., Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology,* 2015. 16(1): p. 224

Rivollat M, et al., Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Science Advances.* 2020. 6(22): p. eaz5344

Rohland N, Hofreiter M, Ancient DNA extraction from bones and teeth. *Nat Protoc.* 2007; 2(7):1756-62. doi: 10.1038/nprot.2007.247. PMID: 17641642

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci.* 2015 Jan 19;370(1660):20130624. doi: 10.1098/rstb.2013.0624. PMID: 25487342; PMCID: PMC4275898

Saupe T, et al., Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Current Biology* 2021. 31(12): p. 2576-2591.e12

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S, Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS ONE.* 2012. 7(3): e34131. doi:10.1371/journal.pone.0034131

Schubert M, Lindgreen S, Orlando L, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes.* 2016 Feb 12; 9:88. doi: 10.1186/s13104-016-1900-2. PMID: 26868221; PMCID: PMC4751634

Scorrano G, et al., Bioarchaeological and palaeogenomic portrait of two Pompeians that died during the eruption of Vesuvius in 79 AD. *Scientific Reports.* 2022. 12. 6468. 10.1038/s41598-022-10899-1

Sirak KA, et al., A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques.* 2017 Jun 1;62(6):283-289. doi: 10.2144/000114558. PMID: 28625158

Sirak K, et al., Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* 2020. 30(3): p. 427-436

- Skoglund P, et al., Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*. 2012. 336(6080): p. 466-9
- Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 2013 Nov; 195(3):693-702. doi: 10.1534/genetics.113.154138. Epub 2013 Sep 11. PMID: 24026093; PMCID: PMC3813857
- van Oven M., PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series*. 2015. 5: p. e392-e394
- Vianello D, et al., HAPLOFIND: a new method for high throughput mtDNA haplogroup assignment. *Hum Mutat*. 2013. 34(9):1189-94
- Weissensteiner H, et al., HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. 2016. 44(W1): p. W58-63
- Willerslev E, et al., Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*. 2003. May 2; 300(5620):791-5. doi: 10.1126/science.1084114. Epub 2003 Apr 17. PMID: 12702808
- Willerslev E, Cooper A, "Ancient DNA." *Proceedings: Biological Sciences*, vol. 272, no. 1558, 2005, pp. 3–16. JSTOR, <http://www.jstor.org/stable/30047500>
- Zalloua PA, et al., Identifying Genetic Traces of Historical Expansions: Phoenician Footprints in the Mediterranean. *The American Journal of Human Genetics*. Volume 83, Issue 5. 2008. Pages 633-642, ISSN 0002-9297
- Zvelebil M, Lillie M, 2000. Transition to agriculture in eastern Europe. In: Price T. editor. *Europe's first farmers*. Cambridge: Cambridge University Press. p 57–92