

DOTTORATO DI RICERCA IN BIOLOGIA

(XXX CICLO)

Coordinatore Prof. Alberto Ugolini

ANALISI GENETICHE DA CALCHI DI POMPEI MEDIANTE *NEXT GENERATION SEQUENCING*

Settore Scientifico Disciplinare BIO/08

Dottorando

Dott.ssa Francesca Amitrano

Tutor

Prof. David Caramelli

Coordinatore

Prof. Alberto Ugolini

Anni 2014/2017

INDICE

1. Introduzione	pag.5
1.1 La valorizzazione dei beni culturali italiani.....	5
1.2 Le analisi biologiche per la valorizzazione dei beni culturali.....	6
1.3 Valorizzazione di reperti scheletrici dell'area archeologica di Pompei.....	7
1.4 L'antica città di Pompei.....	8
1.5 Le analisi molecolari.....	15
2. Scopo del Lavoro	35
3. Materiali e Metodi	37
3.1 Campionamenti.....	37
3.2 Ambiente di lavoro e prevenzione delle contaminazioni.....	44
3.3 Pulizia e polverizzazione dei campioni.....	45
3.4 Estrazione del DNA da polvere d'osso.....	45
3.5 Preparazione delle librerie genomiche <i>half-UDG</i>	48
3.6 Arricchimento delle librerie mediante PCR.....	54
3.7 Cattura del DNA mitocondriale.....	54
3.8 <i>The Next Generation Sequencing Technologies</i> : piattaforme Illumina.....	60
3.9 Sequenziamento mediante piattaforma MiSeq dell'mtDNA catturato.....	64
3.10 Analisi dei dati di sequenza.....	64
3.11 Fasi sperimentali realizzate al Dipartimento di Genetica di Harvard.....	71
4. Risultati	73
4.1 Risultati dell'analisi del DNA mitocondriale.....	76
4.2 Risultati dell'analisi del genoma nucleare.....	80
4.3 Datazioni al radiocarbonio.....	85
5. Discussione	88
6. Bibliografia	91

RIASSUNTO

Grazie alle nuove metodiche di analisi applicate allo studio del DNA antico, quali la cattura delle regioni genomiche endogene d'interesse e il successivo sequenziamento ultramassivo (*high-throughput*) è oggi, possibile, conoscere molte più informazioni riguardanti popoli del passato, talvolta scomparsi in circostanze drammatiche, come il caso degli abitanti dell'antica Pompei all'epoca dell'eruzione del Vesuvio del 79 D.C. I resti scheletrici di molte vittime di quello che si può definire il primo disastro di massa mai documentato nella storia dell'uomo, sono stati ritrovati all'interno di sagome di cenere vulcanica, che avendoli avvolti negli ultimi istanti della loro vita, ne ha preservato per sempre la forma. In epoca recente, la cenere vulcanica è stata sostituita da gesso e cemento liquido a pressione al fine di preservare le sagome (i calchi) contenenti i resti scheletrici delle vittime al proprio interno.

Nel presente lavoro di Dottorato, sedici frammenti ossei e due reperti dentali, recuperati durante il restauro di diciotto calchi di Pompeiani ritrovati in diversi siti dell'area occidentale del sito archeologico di Pompei, sono stati sottoposti ad un'estesa analisi dei genomi mitocondriale e nucleare al fine di caratterizzare per la prima volta dal punto di vista genetico individui residenti nella città di Pompei all'epoca dell'eruzione del 79 D.C.

Le analisi di sequenza effettuate sui genomi mitocondriali hanno dato informazioni circa l'autenticità delle sequenze ottenute e gli aplogruppi di appartenenza. Questi risultati hanno reso possibile chiarire l'assenza di parentela per linea materna tra individui recuperati nello stesso ambiente della così detta Casa del Bracciale d'Oro, sconfessando l'ipotesi precedentemente avanzata dagli Archeologi.

Le analisi di sequenza effettuate sui genomi nucleari, invece, hanno chiarito la determinazione di genere e l'origine geografica dei genomi in esame. Mediante un sequenziamento *shotgun* a bassa copertura realizzato sull'intero genoma nucleare, è stato possibile determinare il sesso degli individui, risultati geneticamente maschi. Inoltre, le frequenze alleliche dei polimorfismi a singolo nucleotide (*SNPs*) catturati dal DNA nucleare e sequenziati, sono risultate molto simili a quelle degli individui del Vicino Oriente, una volta confrontate con quelle presenti in una banca dati di centinaia di individui sia antichi che moderni di Europa ed Asia Occidentale. Questo dato genetico, pertanto, è compatibile con l'ipotesi che i Pompeiani esaminati fossero giunti in questo territorio dal vicino Oriente. Tale convincimento è supportato dalle analisi spettroscopiche dei rapporti isotopici del Piombo.

Infine, campioni esaminati nel presente lavoro, sottoposti a datazione col radiocarbonio, consentono di affermare che si tratta effettivamente di resti antichi di epoca compatibile con l'eruzione del 79 D.C.

1. INTRODUZIONE

1.1. La valorizzazione dei beni culturali Italiani

Nell'ingente patrimonio culturale italiano si riconoscono diverse categorie di beni culturali, che solo per citarne alcune sono rappresentate da beni architettonici, pittorici, sculture, manoscritti, carteggi, libri, oggetti di interesse numismatico, manufatti storici e preistorici e anche reperti paleontologici umani o di animali (figura1).

In conformità con l'articolo 9 della Costituzione (*“La Repubblica promuove lo sviluppo della cultura e la ricerca scientifica e tecnica. Tutela il paesaggio e il patrimonio storico e artistico della Nazione”*), il Codice dei beni culturali e del paesaggio (d.lgs. n. 42 del 2004) ha fissato i concetti guida relativi al pensiero e alle attività rivolte al patrimonio culturale italiano. La tutela è ogni attività diretta a riconoscere, proteggere e conservare un bene del nostro patrimonio culturale, affinché possa essere offerto alla conoscenza e al godimento collettivi; si esplica con il riconoscimento, tramite il procedimento di verifica o dichiarazione dell'interesse culturale di un bene; la protezione e la conservazione. La conservazione è ogni attività svolta allo scopo di mantenere l'integrità, l'identità e l'efficienza funzionale di un bene culturale, in maniera coerente, programmata e coordinata; si esplica con lo studio, inteso come conoscenza approfondita del bene culturale, la prevenzione, la manutenzione e il restauro. Infine, la valorizzazione è ogni attività diretta a *migliorare le condizioni di conoscenza e di conservazione del patrimonio culturale e ad incrementarne la fruizione pubblica, così da trasmettere i valori di cui tale patrimonio è portatore.*

L'interesse istituzionale nei confronti dei beni culturali è rivolto alla loro tutela, conservazione e valorizzazione.

A partire dalla metà degli anni cinquanta del secolo scorso, l'applicazione di metodi scientifici per lo studio dei beni culturali ha registrato un importante sviluppo e al giorno d'oggi le tecniche diagnostiche adottate nell'ambito della conservazione, del restauro e della valorizzazione sono impiegate con approccio multidisciplinare. Le varie branche del settore scientifico mettono a disposizione una numerosa serie di “strumenti” di diagnosi che, grazie alla loro complementarità, consentono di garantire la salvaguardia dei beni, e di *valorizzare, cioè approfondire e promuovere la conoscenza del patrimonio culturale dell'umanità:* tecniche di datazione, microanalisi non distruttive, tecniche geofisiche e microclimatiche, analisi molecolari e isotopiche, permettono di ottenere informazioni estremamente dettagliate per una più completa fruizione del bene, inteso come testimonianza di civiltà.



Figura 1: Alcuni esempi di beni culturali Italiani

1.2. Le analisi biologiche per la valorizzazione di beni culturali

L'avanzamento raggiunto dalle tecnologie scientifiche negli ultimi trent'anni in tutti gli ambiti delle scienze, è stato enorme ed è avvenuto in maniera incredibilmente veloce. Nell'ambito della Biologia, in particolare, lo sviluppo delle procedure e delle tecniche di analisi del patrimonio genetico degli organismi viventi ha permesso di far luce su una moltitudine di aspetti che regolano la vita e le relazioni dei singoli individui e delle collettività a livelli sempre più sofisticati.

Limitatamente all'ambito della genetica umana, il progetto Genoma Umano e gli studi comparativi successivi, hanno messo in evidenza che circa il 99,7% dell'informazione contenuta nel genoma nucleare è identica tra gli individui della specie umana e la variabilità genotipica tra gli individui risiede, in appena lo 0,3% di tutta l'informazione nucleare. L'analisi degli elementi di variabilità interindividuale, noti come polimorfismi del DNA, consente di caratterizzare dal punto di vista molecolare gli individui, stabilendo l'eventuale predisposizione a malattie, la risposta personalizzata ai farmaci, l'identificazione personale in ambito forense, gli accertamenti di parentela per linea paterna. Altamente informativa è anche l'analisi del genoma mitocondriale che rappresenta l'unica analisi genetica possibile quando il genoma nucleare risulta altamente degradato [S.Vai et al., 2015]. Nel caso di resti scheletrici recuperati in siti archeologici, generalmente l'analisi molecolare ancora possibile è quella del DNA mitocondriale che per sua natura e per numero di copie risulta maggiormente

resistente alla degradazione. Tale analisi oltre a stabilire con elevato grado di certezza se i reperti siano da attribuire ad un umano moderno o a specie estinte, ad esempio Neanderthal, consente in casi particolari la localizzazione geografica e l'appartenenza etnica del reperto. Il genoma mitocondriale, inoltre, viene analizzato al fine di ricostruire relazioni parentali per linea materna ed effettuare analisi di popolazione e conoscere le migrazioni dei vari popoli della terra lungo le epoche storiche [D. Caramelli, 2009]. Pertanto, la valorizzazione, ovvero, il consolidamento della conoscenza e la fruizione pubblica di un bene archeologico, viene attualmente realizzata anche attraverso le conoscenze e le tecnologie applicate della Biologia. Queste ultime trovano impiego nel dirimere altresì complicate vicende attributive inerenti opere d'arte, quali i dipinti d'autore [Mascii e Romeo, 2012]. Molte sono le tele non firmate dagli artisti, molte quelle ancora oggi "sospese" in un limbo costruito dagli storici dell'arte; alcune opere addirittura soggette ad attribuzioni diverse da parte dei vari critici d'arte e delle varie scuole di pensiero. Le straordinarie innovazioni in campo scientifico del XX e XXI secolo, adeguatamente applicate, possono offrire un valido contributo nell'attribuzione di un'opera d'arte sottoposta ad indagine. Le analisi fisiche, radiologiche e le indagini al microscopio digitale, ad esempio, hanno portato alla luce in molti dipinti ottocenteschi, la presenza di strati sovrapposti nell'opera, all'interno dei quali risulta intrappolato del materiale di natura biologica, molto probabilmente dell'autore, quali tracce ematiche o altri liquidi corporei e formazioni pilifere, ma anche frammenti di impronte digitali e pollini [Mascii e Romeo, 2012]. Pertanto, il ricorso all'analisi genetica delle tracce biologiche in dipinti di dubbia attribuzione, a seguito del confronto con il profilo genetico del presunto autore, potrebbe fornire indicazioni sull'identità dell'artista.

1.3 Valorizzazione di reperti scheletrici dell'area archeologica di Pompei

Gli scavi archeologici di Pompei hanno portato alla luce i resti dell'antica città di Pompei, insieme a quelli di Ercolano, Stabia ed Oplonti, seppellite da una coltre di ceneri e lapilli durante l'eruzione del Vesuvio del 79 d.C.

Dagli ultimi documenti resi noti dalla Soprintendenza agli scavi, sono stati recuperati inoltre, circa 1150 scheletri umani negli strati di lapilli inferiori, (morti per primi) probabilmente rimasti schiacciati dai crolli degli edifici nella prima fase dell'eruzione; e 650 scheletri recuperati nella parte superiore di cenere vulcanica, morti in un momento successivo a causa delle nubi vulcaniche di ceneri finissime e gas asfissianti che caratterizzarono proprio questo secondo momento dell'eruzione. Le vittime non ebbero il tempo di allontanarsi dall'area circostante il vulcano: vennero sorprese in casa dalla caduta dei solai causata dai violenti

terremoti; altre cercarono invano riparo dai lapilli in depositi, corridoi ed ambienti interni delle Domus dove sono state rinvenute, altre ancora, furono investite da nubi ardenti nelle strade mentre tentavano di fuggire (Figura 2b). Un vero e proprio DISASTRO DI MASSA.

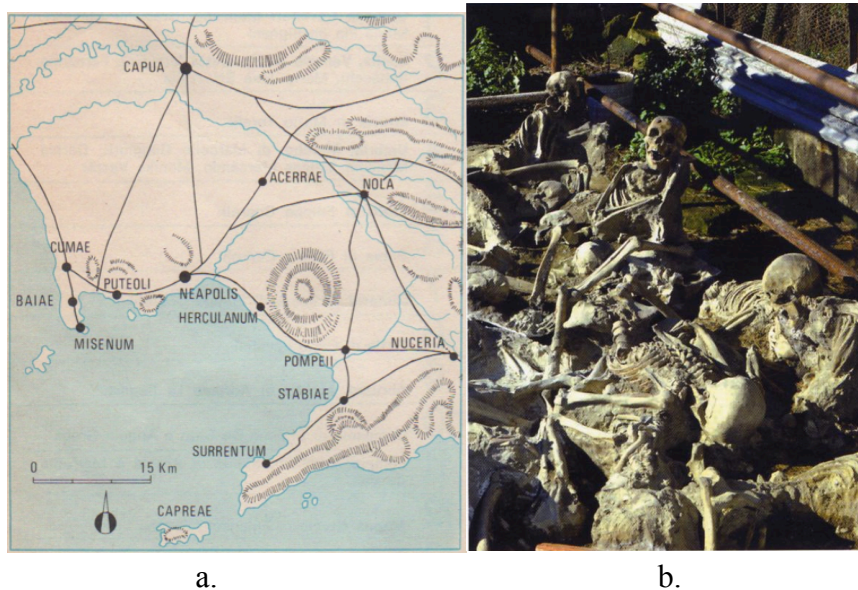


Figura 2: a. litorale dell'agro campano. b. scheletri di fuggiaschi rinvenuti ad Ercolano.

Grazie alle analisi genetiche che oggi si effettuano nel campo dell'Antropologia Molecolare, è possibile conoscere molte più informazioni relative agli abitanti di antiche città, come i Pompeiani ai tempi della grande eruzione. Possiamo rispondere ad interrogativi quali: gli individui recuperati nei medesimi ambienti erano parenti? Di che genere erano? La determinazione genetica del sesso conferma le ipotesi avanzate dagli archeologi? Qual era la composizione etnica della popolazione?

In altre parole, le conoscenze molecolari in aggiunta a quelle archeologiche, consentono di realizzare un'efficace *valorizzazione* di questi reperti effettive testimonianze di civiltà.

1.4 L'antica città di Pompei

1.4.1 Cenni storici: dalle origini alla fine

La storia dell'antica città di Pompei (in latino: *Pompeii*) ha origine nel IX secolo a.C., quando il popolo degli Opici, seppur in forma ancora non stanziale, occupò il territorio posto in posizione strategica, su un pianoro di formazione vulcanica a

circa 30 metri sul livello del mare, alle pendici del Vesuvio (figura 3) ed a breve distanza dalla foce del fiume Sarno, ottima riserva di acqua vista la mancanza di sorgenti in zona. L'assenza di sorgenti o corsi d'acqua sull'altopiano, impedì il suo popolamento nelle epoche più remote, anche se nella vicina valle del Sarno si erano formati alcuni insediamenti, come testimoniano numerose tombe a fossa.

I primi insediamenti stabili risalgono, invece, all' VIII secolo a.C., ad opera degli Osci. Questi fondarono cinque villaggi chiamati *pùmpe*, i quali, intorno al VI secolo a.C., si riunirono in un unico agglomerato cinto di mura. Iniziarono così anche i primi scambi commerciali via mare, con la costruzione di un piccolo porto situato nei pressi della foce del fiume Sarno [R.Etienne,1992].

La fortuna della città fu sin dall'inizio legata alla sua posizione sul mare, che la rendeva il porto dei centri dell'entroterra campano, in concorrenza con le città greche della costa. Naturalmente l'osca Pompei non poteva sottrarsi all'influenza greca, che si estendeva nel golfo di Napoli fino alla penisola sorrentina, includendo anche le isole di Capri e Ischia. L'egemonia greca sulla costa campana venne però ben presto minacciata dall'avanzare prepotente degli Etruschi che conquistarono anche Pompei. Risalgono a questo periodo il Tempio di Apollo e le Terme Stabiane.

Contemporaneamente iniziò anche una lenta ed inarrestabile discesa delle popolazioni sannitiche provenienti dalle zone montane, che conquistarono nel corso del V secolo a.C. tutta la Campania ad eccezione di *Neapolis*, unificandola sotto il proprio dominio.

Pompei dovette subire notevoli trasformazioni urbanistiche ed architettoniche, nel compiere le quali i Sanniti non riuscirono a prescindere dall'influenza greca. E' solo però nel II secolo a.C. col dominio di Roma sulla circolazione delle merci nel Mediterraneo, che la città conobbe un periodo di grande crescita a livello economico, soprattutto attraverso la produzione e l'esportazione di vino ed olio che venivano venduti fino in Provenza ed in Spagna [Etienne R.1992].

Questo stato di benessere si rifletté in un notevolissimo sviluppo dell'edilizia pubblica e privata: furono realizzati il Tempio di Giove e la Basilica nell'area del Foro, a livello privato la dimora signorile della Casa del Fauno. La situazione economica restò florida per molto tempo e furono creati nuovi importanti edifici pubblici, come l'Anfiteatro e l'Odeon.

Un forte impulso alla romanizzazione di Pompei venne dato dall' ascesa dell'imperatore Augusto nel 27 a.C. Le truppe sillane vennero sostituite da quelle augustee, numerosi patrizi portarono in città ricchezze e attività commerciali e si impiantarono nuovi modelli architettonici ed artistici aventi come tema centrale la figura di Augusto. Il Teatro Grande venne restaurato e dedicato ad Augusto, la sacerdotessa Eumachia fece costruire una sorta di nuova basilica chiamata Edificio di Eumachia con statue dedicate alla famiglia augustea, addirittura venne

dedicato un tempio alla Fortuna Augusta. Pompei, ben presto diventò il luogo di villeggiatura del patriziato romano [Devos, 1982].

La città a questo punto si estendeva su quasi 64 ettari e la sua popolazione era di circa 20.000 persone.

Il periodo florido, però, fu bruscamente interrotto il 5 febbraio del 62 d.C., quando un disastroso terremoto, narrato dal filosofo Lucio Anneo Seneca nel sesto libro delle *Naturales quaestiones*, provocò gravissimi danni agli edifici della città. Molti esponenti ricchi si trasferirono in altre zone temendo per la propria incolumità, mentre il commercio calò bruscamente. Gli anni successivi furono investiti nell'imponente opera di ricostruzione che era ancora in atto al momento della fatale eruzione del Vesuvio nel 79 d.C., quando l'antica Pompei fu seppellita completamente da una fitta pioggia di lapilli.

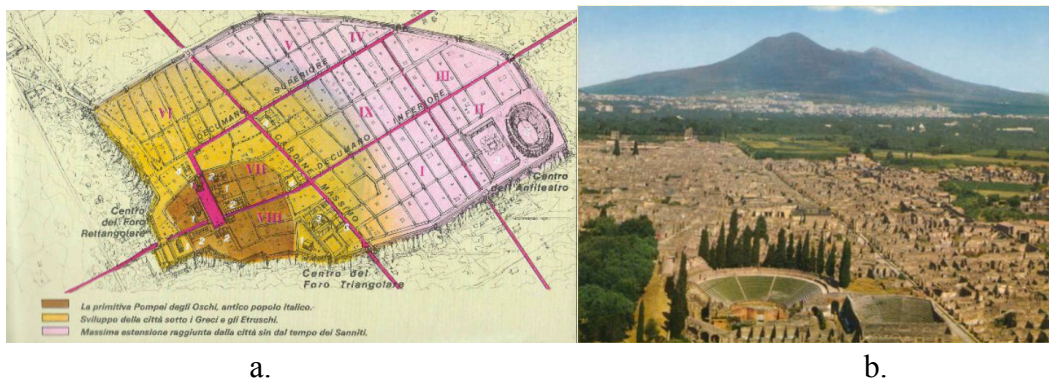


Figura 3: a. schema della mappa di Pompei con indicazione degli ampliamenti susseguitisi. b. fotografia della posizione dell'area dell'antica Pompei rispetto al Vesuvio.

1.4.2 E... la montagna diventò Vesuvio: *cronaca di un disastro di massa*

Alla morte dell'Imperatore Vespasiano, avvenuta il 23 giugno del 79 d.C., suo figlio Tito ereditò un impero florido, dinamico e pacificato. Non poteva immaginare che di lì a poco, su una zona dell'impero, si sarebbe abbattuto il primo disastro di massa documentato nella storia dell'uomo.

L'ammiraglio e scienziato Gaio Plinio Secondo, noto come Plinio il Vecchio, autore della celebre "*Naturalis historia*", nell'estate di quello stesso anno si trovava a Miseno nel golfo di Pozzuoli, al comando della flotta imperiale romana "*Classis Praetoria Misenensis Pia Vindex*". Viveva con una sorella rimasta vedova e con il figlio di lei, Plinio, che per distinguerlo dallo zio venne chiamato "il Giovane". Dalla cima di Capo Miseno, che si erge per oltre 160 metri, si apre la vista sul golfo di Napoli, con il Vesuvio sullo sfondo e le antichissime città

erette sulle sue pendici: Pompei, Ercolano, Stabia ed Oplontis. All'epoca dei fatti al posto di quello che oggi chiamiamo Vesuvio (diventato tale a seguito dell'eruzione del 79 d.C.) sorgeva una montagnola bassa con accanto un'area piana (figura 4): un vulcano inattivo da secoli, ricoperto di boschi e territorio di caccia per i Pompeiani.

Dopo circa otto secoli di inattività, all'alba del 24 Agosto del 79 d. C. (secondo la datazione finora tramandata) venne descritta una nuvola a forma di pino emergere dalla sommità della montagna, seguita nella tarda mattinata, da una violenta esplosione che sotto la spinta dei gas, spaccò il tappo di lava solidificata che ostruiva la fuoriuscita di materiale eruttivo ed andò a frantumarsi in aria. Le masse piroclastiche trasformate così in lapilli e pomici presero a cadere sulla zona a sud-est del vulcano per un raggio di 70 km (figura 5). Su Pompei si depositò un primo strato di lapilli bianchi, di m 1,2, poi un secondo, di m 1,4, di lapilli grigi. La pioggia continuò per quattro giorni, accompagnata da esalazioni di gas tossici asfissianti e cenere. Sopravvennero frequenti scosse di terremoto. Chi non ebbe modo di fuggire, fu soffocato dai gas. Altri restarono schiacciati dai tetti delle case, crollati sotto il peso dei lapilli. Fu una vera tragedia. L'imperatore Tito istituì un'apposita commissione per i soccorsi in Campania. Ma Pompei restò sommersa, anche se nella zona non mancano sepolture e costruzioni fondate sullo strato eruttivo.

Il racconto circostanziato del tragico evento, nelle sue varie fasi, ci è stato tramandato da un testimone oculare d'eccezione, Plinio il Giovane, che, in due lettere indirizzate a Tacito, su richiesta di quest'ultimo, descrive la morte dello zio, Plinio il Vecchio, avvenuta nella seconda fase dell'eruzione sulla spiaggia di Stabia. Il racconto di Plinio il Giovane, in breve, è tramandato come segue [Pliny the younger Letters].

[.....] E' appunto immerso nelle sue letture, Plinio, nella tarda e calda mattinata del 24 agosto 79 d. C., (secondo la datazione maggiormente tramandata fino ad oggi), quando sua sorella lo distoglie per indicargli una strana nube all'orizzonte, verso i monti. L'uomo si alza e va alla ricerca di un punto che consenta una vista migliore di quella nuvola che s'innalza a formare un pino dai colori cangianti fra il bianco e il grigio scuro. Gli è subito chiaro che qualcosa di grave, spaventoso ma anche estremamente interessante dal punto di vista scientifico, sta accadendo, e la sua tipica ostinata curiosità di studioso lo porta subito a superare ogni esitazione e a decidere di salpare per poter osservare più da vicino il fenomeno.

La sua decisione viene ulteriormente rafforzata quando, poco dopo, gli giunge la richiesta di aiuto della nobildonna romana Rectina, moglie di Tasco – entrambi suoi amici – la cui villa sorge ai piedi del Vesuvio. Plinio comprende dunque che si tratta di un'eruzione vulcanica e predispose la sua spedizione che, a questo punto, assume il carattere di soccorso alla gente del posto. Dirige verso Stabia, a

sud di Pompei e dello stesso Vesuvio, per raggiungere il suo amico ed amanuense Pomponiano. Intanto sta calando la sera e sulle navi in avvicinamento piovono già cenere e detriti.

Giunto a casa dell'amico, che trova spaventatissimo tanto per quella pioggia inquietante quanto per le fiamme che la montagna sprigiona in lontananza, lo tranquillizza e, dopo aver cenato, chiede di poter riposare. Poco dopo, però, viene svegliato perché alla cenere si vanno aggiungendo lapilli incandescenti accompagnati da scosse sismiche. L'aria tutt'intorno va facendosi irrespirabile per il diffondersi dell'odore di zolfo e di altre esalazioni.

Plinio si fa accompagnare verso la spiaggia per verificare la possibilità di riprendere il mare, ma quelle esalazioni gli sono fatali: si accascia a terra ed esala l'ultimo respiro.

La gente fugge allontanandosi dagli edifici del centro abitato per paura dei crolli. Viene giorno, ma non c'è luce: quella nuvola oscura il cielo mentre continua ad adagiarsi lentamente per terra. Dovranno passare alcuni giorni prima di ritrovare un po' di tranquillità e quel minimo di lucidità per apprendere che l'eruzione ha cancellato le città di Pompei, Ercolano, Oplontis e Stabia.

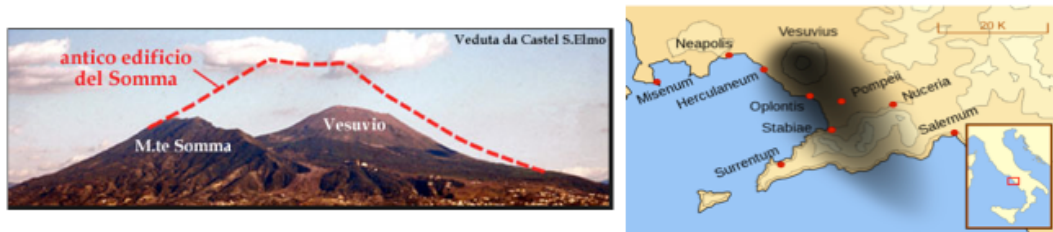


Figura 4: Veduta del vulcano da Castel S. Elmo in Napoli. In rosso l'antico edificio del vulcano prima dell'eruzione. Figura 5: mappa con le aree interessate dall'eruzione.

Da studi recenti sembrerebbe che la data esatta dell'eruzione del 79 d.C non sia il 24 Agosto bensì il 24 Ottobre dello stesso anno. Questo dato è confermato da ritrovamenti archeologici, nonché da osservazioni botaniche e copie delle lettere di Plinio il Giovane in cui è riportata la data dell'eruzione come: *Non[um] Kal[endas] Nove(m)bris*. Nove giorni prima delle calende (il primo giorno di ogni mese nel calendario romano) di Novembre, cioè il 24 Ottobre [Angela A., 2016].

1.4.3 Pompei *riemerge* insieme alle *forme* dei suoi abitanti

Il primo tentativo di scavo che ha interessato l'area dell'antica Pompei si ebbe già qualche anno dopo l'eruzione, sotto l'imperatore Alessandro Severo (222-235

d.C.). Questo scavo, però, fallì a causa dello spessore della coltre di ceneri e lapilli e della vegetazione rigogliosa che nel frattempo aveva ricoperto tutta l'area. Negli anni compresi fra il 1594 e il 1600 iniziarono i primi lavori di scavo, ma la scoperta dei resti dell'antica città fu del tutto casuale nonché misconosciuta. Durante i lavori di costruzione di un canale che trapassava la collina di Pompei per portare le acque del fiume Sarno a Torre Annunziata, furono rinvenuti i resti di edifici, iscrizioni e monete che però non furono subito attribuiti all'antica Pompei ed il violento terremoto del 1631 impedì qualsiasi spunto di ricerca.

I veri e propri scavi dell'antica Pompei iniziarono nel 1738, quando il Re delle Due Sicilie, Carlo di Borbone, affidò la direzione dei lavori all'ingegnere Alcubierre. Si iniziò a scavare in maniera non sistematica, perché il vero intento era quello di trovare oggetti preziosi che potessero arricchire il Museo di Portici. Poi gli scavi furono sospesi perché ad Ercolano fu scoperta la Villa dei Papiri. I lavori ripresero nel 1754 e, dopo un solo anno di scavo, fu rinvenuta la Villa di Giulia Felice, la Porta di Ercolano ed un'epigrafe.

Con la salita al potere di Ferdinando I delle Due Sicilie, nel periodo compreso tra il 1759 e 1799 fu riportata alla luce parte della città, questa volta non più riseppellita ma rimasta a vista, grazie anche a un sistema di scavo sistematico, voluto dal direttore Francisco la Vega, il quale preferiva che i reperti, soprattutto gli affreschi parietali, rimanessero ai muri e non asportati per essere esposti altrove. Negli anni 1760-1772 l'attenzione si spostò nella zona nord-occidentale della città, con le esplorazioni della Villa di Diomede, della Casa del Chirurgo e della Via dei Sepolcri, dove furono rinvenuti oltre a monete d'oro ed argento, anche diciotto corpi di vittime dell'eruzione.

Con Giuseppe Bonaparte Re di Napoli, agli inizi del 1800, gli scavi ebbero un nuovo incentivo: vennero alla luce la strada tra la Villa di Diomede e la Casa di Sallustio, le case dette del Poeta tragico e del Fauno, il Foro e la "Basilica" nonché venne individuata la cinta muraria della città.

E' in questi anni che Pompei diventava una sorta di "museo all'aperto"; furono rinvenuti il Tempio della Fortuna Augusta e le Terme del Foro.

A seguito dell'unità d'Italia (1861) la direzione degli scavi fu affidata al genio dell'archeologo napoletano Giuseppe Fiorelli (figura 6). Questi fu il primo a dare un'organizzazione sistematica alle opere di scavo: divise la città in *regiones*, cioè quartieri, ed in *insulae*, ovvero agglomerati di case (figura 7).

Al genio di Fiorelli si deve anche un'altra importante innovazione: l'introduzione della tecnica del calco in gesso per ridare forma alle vittime dell'eruzione, restituendo l'istantanea dei loro ultimi momenti di vita (figura 8).

A Pompei esistono due tipi di vittime: chi è morto per i crolli dovuti al peso della "grandine vulcanica" o dei terremoti, e chi è stato ucciso dalla nube vulcanica di cenere finissima incandescente e gas solforici asfissianti. In questo secondo caso,

oltre allo scheletro c'è anche l'impronta lasciata dal corpo della vittima nella cenere compatta. Presupposto questo che ha reso possibile realizzarne il calco, inizialmente in gesso e poi in cemento liquido a pressione (2 atm). Con il tempo la decomposizione del materiale organico ha lasciato solo lo scheletro creando il vuoto nella gabbia di cenere indurita intorno al corpo e di cui ha mantenuto perfettamente la forma. Pertanto, riempiendo il vuoto con gesso o cemento liquido a pressione e lasciando seccare, si ottengono le esatte fattezze della vittima nei suoi ultimi momenti di vita, dopo quasi duemila anni.



Figura 6 : Giuseppe Fiorelli



Figura 7: Mappa della città di Pompei con la suddivisione in *regiones* (in numeri romani) realizzata da G.Fiorelli. I numeri nei cerchi rossi indicano le porte dell'antica città.



Figura 8: Fotografie dei calchi delle vittime ritrovate nell'orto dei Fuggiaschi a Pompei.

A Fiorelli seguirono altri direttori che perfezionarono la tecnica del recupero e della conservazione degli edifici. Sotto la direzione di Amedeo Maiuri furono riportati alla luce edifici del calibro della Villa dei Misteri e si iniziò a scavare anche negli strati sottostanti al livello del 79 d.C. per riscoprire la storia più antica di Pompei. Successivamente, con la direzione di Alfonso De Franciscis, ci fu un rallentamento degli scavi in quanto il sito archeologico avendo raggiunto ormai dimensioni notevoli, necessitava di continue opere di manutenzione. Pertanto, furono accantonate ulteriori opere di scavo per incentivare il restauro e la manutenzione delle opere già portate alla luce.

Nel 1980, poi, il terremoto dell'Irpinia rallentò per lungo tempo i lavori di scavo, che ripresero successivamente grazie ai finanziamenti del Fondo Investimenti e Occupazione, utilizzati nel "Progetto Pompei" ai fini della valorizzazione dell'area archeologica.

Quest'ultima, venne dichiarata dall'UNESCO patrimonio dell'umanità nel 1997, e nominata dalla commissione stessa "senza uguali al mondo".

Nel 2012 entrò in vigore, sotto il nome di "Grande Progetto Pompei", un intervento rilevante ed economicamente impegnativo volto alla riqualificazione del sito archeologico di Pompei entro il 2015. Il progetto frutto di un'intesa inter-istituzionale voluta dal Governo Italiano aveva l'obiettivo di ridurre il rischio idrogeologico con la messa in sicurezza dei terrapieni non scavati; consolidare e restaurare le murature e le superfici decorate; migliorare la protezione degli edifici dalle intemperie con conseguente aumento delle aree visitabili; rafforzare il sistema di videosorveglianza.

1.5 Le Analisi molecolari

1.5.1 Il DNA antico (*aDNA*)

Con il termine **DNA antico** (**aDNA** = *ancient DNA*) è definita qualsiasi traccia di DNA proveniente da un organismo morto o da una parte di esso o da campioni biologici non recenti. Questo tipo di DNA può, quindi, essere estratto da un'ampia varietà di materiali biologici, diversi per origine, età e stato di conservazione che prevalentemente sono rappresentati da ossa, denti, formazioni pilifere, resti mummificati animali e vegetali e coproliti.

Il primo studio di ciò che sarebbe stato chiamato **aDNA** uscì nel 1984, quando Russ Higuchi e i suoi colleghi all'università di Berkeley, riuscirono ad estrarre e sequenziare tracce di DNA proveniente da un campione museale di quagga (*Equus quagga quagga*) una sottospecie estinta della Zebra delle pianure che

un tempo viveva in Sudafrica [Higuchi R.,1984]. Negli anni successivi, il recupero di sequenze genomiche da mummie egizie, confermò la possibilità che il DNA poteva essere estratto e analizzato con successo anche da resti antichi di parecchie migliaia di anni [Pääbo, 1985]. Ciononostante, i laboriosi processi che erano richiesti per sequenziare tale DNA (attraverso la clonazione batterica) erano un freno effettivo allo sviluppo delle ricerche sul DNA antico. Soltanto dal 1986 in poi, con l'introduzione della reazione a catena della polimerasi (*Polymerase Chain Reaction*, PCR) ad opera di Kary Mullis, il campo di ricerca sul DNA antico ha avuto la possibilità di progredire rapidamente. L'invenzione della reazione a catena della polimerasi ha rappresentato una vera e propria rivoluzione nell'ambito dell'intera Biologia Molecolare. Nel campo dell'Antropologia Molecolare ha facilitato notevolmente l'analisi del DNA antico. Questa tecnica consente di replicare selettivamente, in un sistema acellulare, un singolo frammento di DNA d'interesse sulla base della sua sequenza nucleotidica, quantunque esso sia disperso in una moltitudine di altri frammenti di DNA, condizione normale per il DNA antico. Dalla fine degli anni '80 in poi gli sviluppi nel campo del DNA antico hanno consentito di indirizzare importanti interrogativi dell'Antropologia classica. Accertare i rapporti di parentela nel caso di sepolture multiple (per linea materna attraverso l'analisi del DNA mitocondriale, e per linea paterna attraverso l'analisi del cromosoma Y); ricostruire i flussi migratori delle popolazioni umane; identificazione personale e di genere; ricostruire la filogenesi di specie estinte e calcolare la distanza evolutiva da quelle moderne [Caramelli D., Lalueza-Fox et al., 2003]; studiare le paleopatologie (diffusione di agenti infettivi correlati a fattori socioeconomici o patologie geneticamente trasmissibili rilevanti a livello popolazionistico).

Lo studio del DNA antico ha consentito anche di identificare geni chiave nell'evoluzione dell'uomo. E' il caso del gene nucleare *FOXP2* (7q31) implicato nello sviluppo del linguaggio articolato dell'uomo moderno. Nella nostra specie, il fattore di trascrizione codificato da questo gene mostra due sostituzioni aminoacidiche fondamentali per lo sviluppo del linguaggio articolato. Queste sono presenti anche nel gene *FOXP2* di Neanderthal, mentre sono assenti nello scimpanzè [Krause et al., 2007]. Molto recentemente, analisi di sequenziamento *ultramassivo* dei genomi di 5 campioni estratti da placche dentali calcificate di Neandertaliani recuperati in siti diversi, hanno consentito di ricavare informazioni relative alla dieta, al comportamento ed alle malattie di questi ominidi [Laura S. Weyrich et al., 2017].

1.5.2 Le problematiche dell'analisi dell'*aDNA*

La manipolazione, l'estrazione e l'analisi del DNA antico, come pure l'interpretazione dei risultati, sono caratterizzate da una serie di problemi causati principalmente dai tre motivi sottoelencati:

- ❖ la degradazione della doppia elica antica;
- ❖ il basso numero di copie in cui è presente (*Low Copy Number*);
- ❖ l'elevato grado di contaminazione.

1.5.2.1 La degradazione della doppia elica antica e marcatori di autenticità

La doppia elica del DNA, nella forma idratata in cui si trova nelle cellule è una molecola relativamente stabile grazie ai molteplici meccanismi di riparo di cui è dotata la cellula vivente. Quando un organismo muore tali meccanismi di riparo vengono meno e quindi il DNA, come tutte le altre molecole di cui è composta la cellula, va incontro ad una serie di alterazioni biochimiche, indicate nell'insieme come *diagenesi*, che culminano nella sua completa degradazione. I fattori responsabili di tale degradazione sono sia di tipo endogeno, propri dell'organismo morto, sia esogeni cioè caratteristici dell'ambiente in cui l'organismo si trova.

Tra i fattori endogeni, un ruolo fondamentale è rivestito dagli enzimi lisosomiali che vengono liberati nella cellula morta ove operano azioni degradative di tutte le molecole biologiche. Alla degradazione del DNA cooperano, in particolare proteasi e nucleasi lisosomiali. Le prime idrolizzando le proteine che impacchettano la struttura cromatinica, rendono il DNA facilmente accessibile alle nucleasi endogene e microbiche eventualmente presenti. Questi enzimi sono in grado di idrolizzare i legami fosfodiesterici (5'fosfato-3'OH) che uniscono i monomeri nucleotidici a formare il polimero di acido nucleico il quale pertanto è ridotto in frammenti. **E' stato constatato infatti che il *aDNA* è sempre presente sotto forma di frammenti la cui lunghezza è compresa tra 40 e 500 bp** [Dabney et al. 2013]. L'altra azione degradativa operata dalle nucleasi è l'idrolisi del legame covalente N-beta-glicosidico tra la base azotata (purina o pirimidina) e lo zucchero 2'-desossiribosio, con creazione di un sito abasico (AP, a-purinico, a-pirimidinico).

E' stato osservato che la velocità con cui le basi puriniche (adenina e guanina) vengono liberate è circa 20 volte superiore a quella delle basi pirimidiniche, (timina e citosina) pertanto l'idrolisi delle basi è indicata generalmente come "depurinazione". Questa reazione è influenzata dal pH del mezzo: il tasso di depurinazione diminuisce all'aumentare del pH fino al valore di pH 6.0, oltre il

quale tale correlazione va annullandosi. Processi fermentativi anaerobici dei microrganismi del suolo contribuiscono ad abbassare il pH favorendo la degradazione del DNA [Gill-King,1996]. La presenza di siti abasici, porta inoltre alla formazione di addotti tra il 2'-desossiribosio del sito abasico e il gruppo amminico della base sul filamento opposto, e tra DNA e proteine [Sancar et.al 2004].

Un'altra importante reazione di idrolisi che subisce l'aDNA è rappresentata dalla deaminazione delle basi azotate. In particolare le basi con gruppi amminici secondari (adenina, citosina, 5-metilcitosina e guanina) subiscono la perdita, anche spontanea, di questi gruppi -NH₂, con conseguenti modificazioni delle basi stesse. Più frequentemente si osserva la transizione da citosina a uracile (U), un analogo della Timina (T); e da adenina a ipoxantina (HX), un analogo della guanina (G). Per semplicità si parla di transizioni C->T ed A->G, rispettivamente [Hofreiter et al.,2001]. Queste modifiche, in fase di analisi in particolare durante la PCR portano ad un'errata lettura della sequenza dello stampo da parte della DNA polimerasi che quindi produce un'errata incorporazione delle basi durante l'estensione del filamento di nuova sintesi. Pertanto, si dice che la deaminazione delle basi è responsabile delle lesioni da *miscoding*, ***misincorporazioni causate dalle caratteristiche precipue dell'elica stampo antica***. Bisogna però distinguere le misincorporazioni dovute all'elica stampo antica e quindi in altri termini prove di autenticità di aDNA, dalle misincorporazioni dovute agli errori della DNA polimerasi che possono avvenire in qualsiasi PCR a prescindere dall'integrità della molecola stampo di partenza. Ebbene, nel 2007 Brotherton et al. dimostrarono che **soltanto la deaminazione della citosina ad uracile è il danno postmortem che identifica come autentica una molecola antica** [Brotherton et al. 2007] (figura 9). Le transizioni C->T poi, caratteristiche delle molecole antiche risultano maggiormente concentrate all'estremità 5' e quindi le quelle G-> A al 3' delle molecole antiche [Briggs AW et al. 2007; Krause et al. 2010].

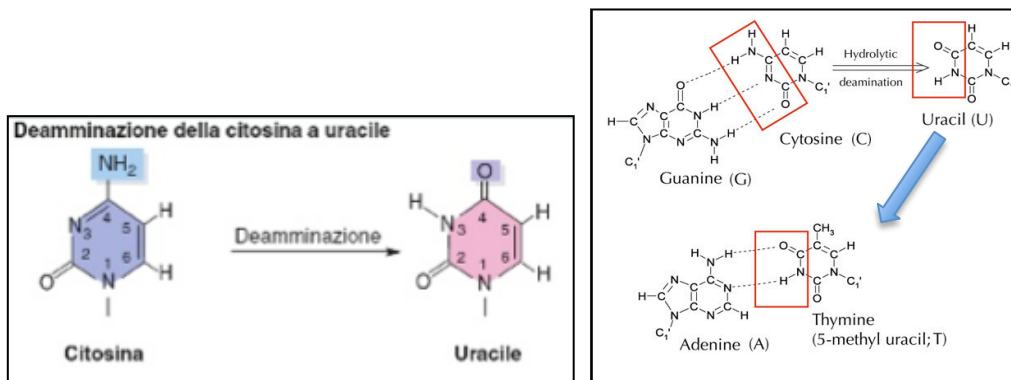


Figura 9: Deaminazione della Citosina ad Uracile e successiva incorporazione della base Timina.

Anche i fattori ambientali influenzano notevolmente lo stato di degradazione di un DNA antico. La presenza di acqua favorisce la crescita microbica e quindi l'attività idrolitica degli enzimi microbici è favorita. Un DNA recuperato da tessuti essiccati, al contrario, risulta protetto dagli attacchi microbici, ma maggiormente suscettibile ai danni di tipo ossidativo operati dalle specie reattive dell'ossigeno [Paabo 1989]. La temperatura ambientale è un altro parametro da considerare. Le basse temperature preservano i resti antichi per lunghi periodi di tempo: 50.000 anni nelle calotte di ghiaccio e un tempo anche superiore nel permafrost. Anche in queste condizioni però i danni al DNA antico sono evidenti e fondamentalmente connessi agli eventi endogeni che hanno luogo dopo la morte, all'esposizione alle radiazioni UV e ai radicali liberi dell'ossigeno (in assenza di sistemi antiossidanti attivi).

In linea generale, dunque, una molecola di aDNA ha poche probabilità di conservarsi in un ambiente caldo, umido, in presenza di ossigeno e microrganismi. Al contrario, le probabilità di conservazione aumentano se il sito presenta condizioni ostili allo sviluppo dei microrganismi, basse temperature, aridità (brughiera), alto tenore di salinità ed assenza di ossigeno.

1.5.2.2 Il basso numero di copie dell' aDNA (*Low Copy Number*)

Un altro problema per l'analisi del DNA antico o altamente degradato, è il basso numero di copie (*Low Copy Number*) con cui la doppia elica antica è presente nel campione stesso. Per basso numero di copie, ci si riferisce ad una quantità di DNA estratta dal campione, inferiore a 100 pg. Con l'introduzione della *Polymerase Chain Reaction* (PCR) gli ostacoli all'analisi di un tale DNA sono stati, almeno in parte, superati. In effetti con la PCR è possibile *arricchire* il numero di copie in cui un dato frammento di DNA d'interesse è presente nel campione, partendo virtualmente anche da una singola molecola, per riprodurre in breve tempo milioni di copie del frammento. L'interpretazione dei risultati ottenuti in tali condizioni, nota come *low copy number typing* (*LCN typing*), ha però non poche criticità. Infatti, quando sono presenti poche molecole di DNA stampo, in fase di amplificazione possono avvenire errori stocastici che portano a perdita allelica o a squilibrio (*imbalance*) fra i due alleli di una coppia di omologhi. Molte sono le strategie messe in atto nei diversi laboratori per affrontare il problema del *LCN typing*, ma non c'è ancora una strategia condivisa che non presenti difficoltà interpretative [Butler John M. 2011]. Una possibilità ad esempio è aumentare il numero di cicli della PCR per migliorare la resa dell'amplificazione, però anche in questo caso oltre agli errori stocastici menzionati, aumenta la probabilità di amplificare DNA contaminante,

specialmente nel caso in cui si lavori con DNA antico umano [Butler John M. 2011].

1.5.2.3 aDNA: contaminazione e criteri di autenticità nell'analisi molecolare

Un resto antico nel corso della sua storia tafonomica va inevitabilmente incontro a contaminazioni. In primo luogo, i microrganismi saprofiti che crescono a spese della materia organica in decomposizione ne degradano le componenti cellulari impoverendola del DNA originale mentre la arricchiscono del proprio materiale genetico. Questo tipo di contaminazione è facilmente risolta nell'analisi molecolare: la PCR consente di selezionare ed amplificare il DNA originale proprio del campione discriminandolo dal contaminante di origine microbica. Il tipo di contaminazione più difficile da risolvere nel caso dell'analisi del DNA di un resto umano antico, è la contaminazione da parte di DNA umano moderno. Un reperto infatti, subisce una serie di manipolazioni ad opera dell'archeologo, del personale del museo, dello studioso di morfologia, ed infine dell'archeologo molecolare che tendono ad inquinare il reperto col proprio DNA derivante, ad esempio, dalle cellule epidermiche desquamate. Pertanto, prima di procedere all'analisi molecolare, risulta utile, per quanto possibile, ottenere il maggior numero di informazioni circa il sito in cui il reperto è stato recuperato, le modalità di recupero e conservazione del reperto stesso, nonché il numero di persone che ne sono venute a contatto. Il materiale genetico che si può estrarre da un reperto umano antico è dunque, rappresentato da una miscela di DNA di cui la quota maggioritaria è rappresentata da DNA contaminante di diversa origine (microbica, umana moderna) e quella minoritaria da DNA endogeno del campione antico degradato in maniera più o meno spinta. L'analisi della lunghezza dei frammenti amplificati dal campione, può dare un'indicazione circa l'autenticità delle sequenze in esame. Un'analisi ad esempio, del DNA nucleare di mummie egizie rivelava la presenza di frammenti lunghi 3,4 kb [Paabo,1985]. Successivamente, tali sequenze insolitamente lunghe per essere DNA antico altamente degradato, si sono rivelate per lo più DNA contaminante [Pilli et al.2016]. Oltre alla presenza di DNA umano moderno, un'ulteriore fonte di contaminazione frequente nell'analisi del DNA antico è rappresentata dalle molecole di DNA già amplificate che potrebbero essere presenti in elevata concentrazione in soluzioni e reagenti precedentemente utilizzati nelle fasi sperimentali. E' necessario, quindi, tenere sotto controllo l'inquinamento del laboratorio che tratta *aDNA*, attuando alcuni accorgimenti. Durante le fasi analitiche, ad esempio, bisogna indossare guanti, mascherine, cuffie, e camici sterili; aliquotare sterilmente tutti i reagenti utilizzati nelle fasi sperimentali in quantità necessarie a lavorare sul singolo

campione al fine di evitarne l'uso ripetuto. Tipizzare il DNA degli operatori in un apposito laboratorio in cui vengono eseguite analisi sul DNA moderno, in modo tale da poterlo identificare qualora venisse accidentalmente amplificato insieme o al posto del DNA antico.

In un articolo pubblicato sulla rivista "*Science*" nel 2000 gli autori *Cooper A. e Poinar HN*, hanno definito i criteri che devono essere soddisfatti per realizzare uno studio sul DNA antico e che vengono comunemente indicati come "Golden Criteria" [Cooper A. Poinar HN 2000].

- Valutare il grado di preservazione del campione antico mediante opportune analisi biochimiche, in maniera tale da avere già una stima dei reperti che contengono DNA meglio conservato, in accordo con il loro grado di conservazione biochimica (collagene, racemizzazione degli aminoacidi).
- Analizzare campioni recuperati dallo stesso sito del reperto in esame, allo scopo di ottenere informazioni riguardo le contaminazioni ambientali.
- Organizzare il laboratorio in spazi fisicamente separati per le fasi di pre-amplificazione e post-amplificazione.
- Allestire controlli negativi in tutte le fasi analitiche, in modo da monitorare lo stato dei reagenti, l'assenza cioè di contaminazione nelle soluzioni utilizzate per le reazioni.
- Allestire differenti amplificazioni dallo stesso o da differenti estratti dello stesso campione. Questo al fine di controllare le contaminazioni di un'estrazione o di un'amplificazione; verificare le misincorporazioni.
- E' indispensabile quantificare il DNA estratto per stabilire se e come procedere alle successive fasi di amplificazione.
- Ai fini della riproducibilità del dato: i campioni devono essere analizzati e i risultati devono essere riprodotti in laboratori diversi. In questo modo vengono controllati eventuali contaminanti che possono essere eventualmente presenti in un laboratorio e non in un altro.
- Data la degradazione che subisce il DNA antico, i frammenti che vengono amplificati devono avere generalmente dimensioni di circa 200bp. Nell'analisi dei reperti antichi, infatti, si osserva una correlazione inversa fra l'efficienza di amplificazione e la lunghezza dell'amplificato stesso [Paabo et al. 1989]. Anche se nel caso di resti recuperati dal permafrost sono risultati autentici anche frammenti di 1,6 Kb [Caramelli 2009].

Nonostante si osservino scrupolosamente i criteri citati, è sempre presente la possibilità di analizzare DNA contaminate insieme o al posto del DNA endogeno. Questo è frequente soprattutto in caso di reperti umani antichi contaminati da DNA umano moderno prima delle fasi molecolari.

1.5.3 Marcatori dell'analisi del DNA antico

1.5.3.1 I polimorfismi del DNA nucleare

Il genoma nucleare umano è costituito in assetto diploide (2N) da 6,6 Gb (gigabasi) organizzate in 23 coppie di cromosomi (22 coppie di cromosomi omologhi autosomi, ed una coppia di eterosomi X ed Y: nella donna gli eterosomi sono rappresentati da due copie del cromosoma X; nell'uomo da una copia del cromosoma X ed una copia del cromosoma Y) (figura 10). Per ciascuna coppia, un cromosoma è ereditato dalla madre e l'altro dal padre. Nella specie umana il sesso eterogametico è il maschio, pertanto, un individuo sarà geneticamente maschio se lo zigote (2N) da cui ha avuto origine era formato dalla fecondazione della cellula uovo (N) da parte di uno spermatozoo (N) recante il cromosoma Y come eterosoma. Poiché il cromosoma Y è trasmesso in una famiglia esclusivamente dal padre ai figli maschi, è possibile verificare relazioni di parentela per linea paterna proprio analizzando la trasmissione del cromosoma Y.

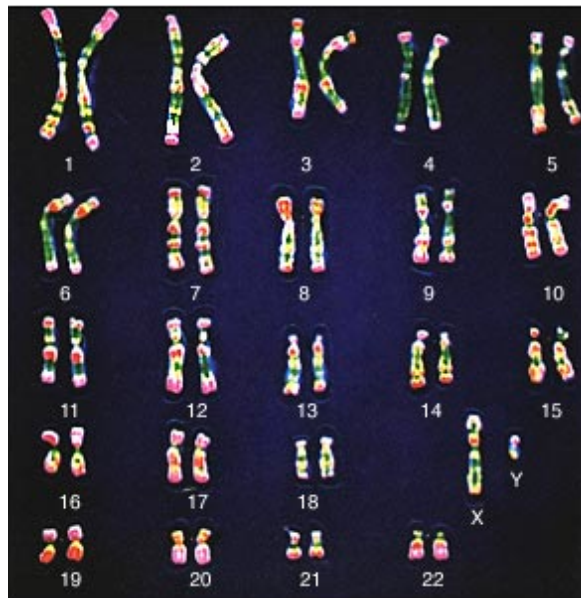


Figura 10 : i 46 cromosomi umani (individuo di genere maschile)

Tra le caratteristiche del genoma nucleare umano, l'elevata presenza di sequenze ripetute per il 90% localizzate in regioni inter geniche e le differenze interindividuali contenute in appena lo 0,3% del genoma nucleare totale, sono state ampiamente utilizzate per studi di popolazione nell'ambito

dell'Antropologia Molecolare e per l'identificazione personale in ambito forense [Stoneking M.1997; Li R. 2015].

Le differenze tra gli individui a livello del DNA, sono indicate con il nome generico di polimorfismi del DNA. Tra questi possiamo distinguere fondamentalmente polimorfismi di lunghezza; polimorfismi di sequenza; inserzioni e delezioni (InDel); variazioni del numero di copie (Copy Number Variation, *CNV*) (figura 11).

I polimorfismi maggiormente utilizzati fino ad oggi, sono rappresentati dai polimorfismi di lunghezza *Short Tandem Repeat (STR)*: sequenze di DNA ripetute in tandem in cui l'unità di ripetizione può essere costituita dalle 2 alle 6 bp. Tra gli STR, quelli più utilizzati sono i tetranucleotidi (unità di ripetizione costituita da 4 basi) per diversi motivi: sono molto comuni nel genoma, i loro prodotti di amplificazione mostrano una bassa percentuale di artefatti. Generalmente i loci *STR* utilizzati sono anche altamente polimorfici nella popolazione (da 5 a 20 alleli per marcatore), pertanto, sono molto utilizzati per l'identificazione personale in ambito forense [Butler J. 2011]. Le dimensioni dei loro prodotti di amplificazione, comprese tra 100-400 bp, però, risultano abbastanza elevate nel caso di molecole di DNA stampo altamente degradate [Butler 2011]. Wiegand e Kleiber nel 2001 dimostrarono che loci *STR* di campioni di DNA degradato potevano essere tipizzati con successo utilizzando *primers* di PCR ridisegnanti per produrre ampliconi di dimensioni inferiori a 125 bp, i "*miniSTR*" [Wiegand e Kleiber nel 2001].

La vera promessa nell'analisi del DNA antico altamente degradato, è rappresentata però, dall'utilizzo dei polimorfismi di sequenza a singolo nucleotide (*SNPs*). Si tratta cioè di loci le cui forme alleliche differiscono nella sequenza nucleotidica per una singola base. I loci *SNP* sono per la maggioranza biallelici nella popolazione: per un dato SNP c'è l'allele comune e l'allele variante, con la variazione di una base avente una frequenza superiore all'1% nella popolazione. Gli *SNP* offrono importanti vantaggi rispetto agli *STR* e ai *miniSTR* ai fini dell'analisi del DNA antico. In primo luogo, sono più abbondanti nel genoma umano (ce n'è 1 ogni 1000 bp), i prodotti di amplificazione risultano notevolmente corti, meno di 100bp, caratteristica che aumenta notevolmente l'efficienza dell'analisi se lo stampo è una molecola di DNA altamente degradata (figura 12). Grazie alle attuali metodiche di *arricchimento* di un DNA d'interesse all'interno della miscela di DNA presente nel campione, in aggiunta alle nuove tecnologie di sequenziamento ultramassivo, il sequenziamento degli *SNPs* potrà nel prossimo futuro rappresentare l'analisi di elezione per la tipizzazione dei campioni antichi.

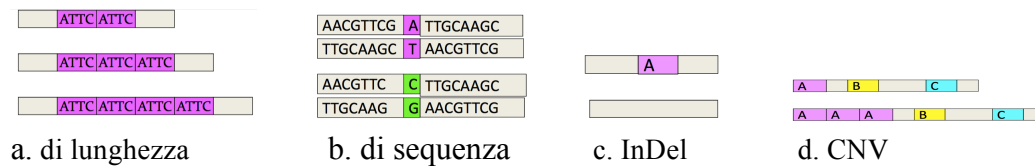


Figura 11: Polimorfismi del DNA.

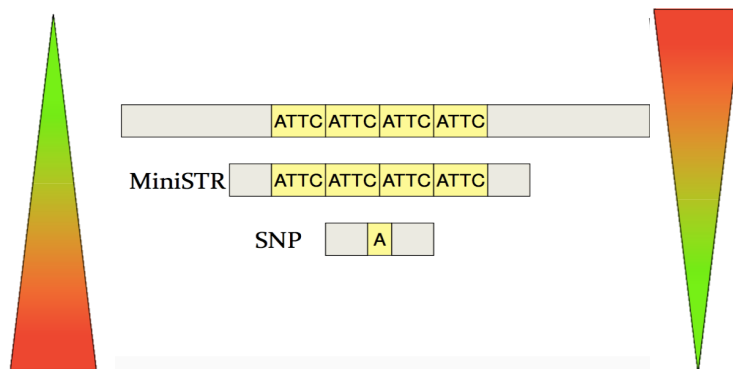


Figura 12: Schema che mostra la relazione inversa tra l'efficienza di analisi del DNA degradato (freccia a sinistra) e la taglia dei prodotti di amplificazione (freccia a destra).

Un altro tipo di polimorfismo biallelico nella specie umana è rappresentato dalle Inserzioni Delezioni (InDel). Un InDel può essere l'inserzione o la delezione di un segmento di DNA che va da un nucleotide a centinaia di nucleotidi. Dato un locus InDel, i due alleli sono semplicemente classificati come allele lungo e allele corto in base alla presenza o assenza dell'inserzione. Un esempio molto noto di questo tipo di polimorfismo è rappresentato dagli elementi *Alu*, sequenze della famiglia *SINE* (*Short Interspersed Nuclear Elements*) una forma di DNA ripetuto lunga meno di 500 bp e dislocata in vari punti del genoma. Le *SINE* più comuni nei primati e nell'uomo ed anche meglio studiate sono le sequenze *Alu*, i polimorfismi InDel. Questi elementi *Alu* sono sequenze lunghe circa 300 bp la cui presenza nel genoma viene individuata dal fatto che contengono il sito di restrizione dell'enzima endonucleasi *AluI*, da qui il nome. Il genoma umano, inoltre, contiene oltre 1 milione di copie di sequenze *Alu*, parziali o complete, disperse tra i geni all'interno degli introni, costituendo l'11% circa del genoma nucleare umano [Caramelli 2009]. L'inserzione di un elemento *Alu* ad un particolare locus può essere considerata come un evento unico. Una volta inseriti, gli elementi *Alu* sono marcatori genetici stabili, non sono sottoposti a perdita o a riarrangiamenti [Butler 2009]. Metodi identificativi di lunga data (PCR, elettroforesi su gel d'agarosio e visualizzazione mediante bromuro d'etidio), riescono a mettere in evidenza le inserzioni *Alu* specifiche del genoma umano. La presenza di un'inserzione *Alu* è indicata da un prodotto di PCR di 400 bp, mentre

l'assenza dell'inserzione risulterà in un amplicone di 100bp (figura 13). Poiché molte sequenze *Alu* sono uniche per gli umani, è possibile utilizzarle per l'identificazione di DNA umano [Butler 2009].

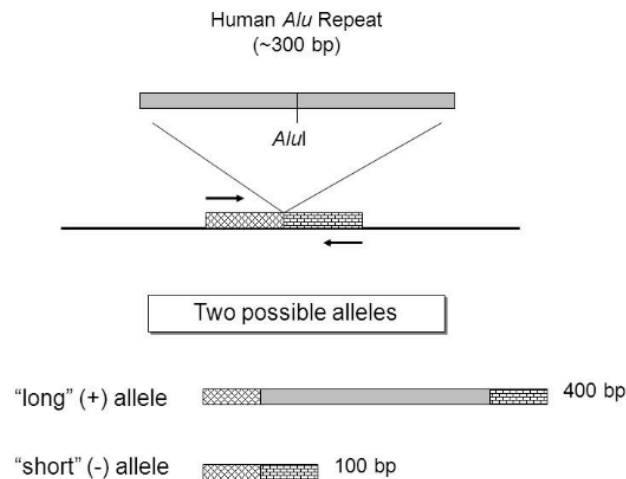


Figura 13: Rappresentazione schematica della rivelazione per PCR del polimorfismo InDel Alu ad una data posizione nel genoma umano, [Butler 2009].

1.5.3.2 Applicazioni dell'analisi degli *SNPs*

I polimorfismi a singolo nucleotide, *SNPs*, del genoma umano possono essere analizzati fondamentalmente per 4 diversi scopi: 1) identificazione personale [Budowle e van Daal 2008]; 2) identificazione dell'origine geografica di un genoma; 3) identificazione di discendenza; 4) indicazioni sulle caratteristiche fenotipiche (colore degli occhi, della pelle, dei capelli) [Butler 2011]. *SNPs* con caratteristiche diverse sono impiegati per scopi diversi (tabella 1). Per l'origine geografica di un genoma, sono stati selezionati *SNPs* che mostrano una bassa eterozigosità all'interno di una stessa etnia ed un'alta variabilità tra le etnie, ad es. *SNPs* molto rari negli africani e molto frequenti negli europei. *SNPs* di questo tipo hanno anche un tasso di mutazione molto basso e pertanto sono considerabili "fissati" in una popolazione. Pertanto, sia questo tipo di *SNPs* che le inserzioni *Alu* sono spesso trovati etnia specifici [Butler 2009]. Nel caso dei caratteri fenotipici, es. colore degli occhi dei capelli etc., si tratta di caratteri multigenici e multifattoriali in cui è coinvolta anche la componente ambientale. Ciononostante in determinati geni sono stati identificati *SNPs* correlati con un determinato fenotipo, nel senso che la presenza di una variante o dell'altra dello *SNPs* aumenta le probabilità di riscontrare un dato fenotipo rispetto all'altro [Butler 2011].

Categoria <i>SNPs</i>	Caratteristiche
Marcatori identificativi	<i>SNPs</i> non associati che analizzati in <i>multiplex</i> , mostrano una probabilità molto bassa che due individui abbiano per essi un identico genotipo multi-locus.
Marcatori di discendenza	<i>SNPs</i> in loci strettamente associati e multiallelici nella popolazione, che analizzati in <i>multiplex</i> funzionano come marcatori di parentela
Marcatori etnici	<i>SNPs</i> che collettivamente indicano con alta probabilità l'origine geografica di un individuo
Marcatori fenotipici	<i>SNPs</i> che indicano con alta probabilità tratti fenotipici di un individuo: colore pelle, capelli, occhi.

Tabella 1: Categorie dei marcatori *SNPs* [Butler 2011].

1.5.3.3 La prima identificazione di genere su reperti ossei dell'antica Pompei

La presenza di DNA endogeno ancora analizzabile nei resti scheletrici delle vittime dell'eruzione del 79 D.C. a Pompei, è stata per lungo tempo messa in dubbio a causa delle caratteristiche dell'ambiente di sepoltura e delle modalità della morte. L'analisi genetica di tali reperti è però di eccezionale valore scientifico proprio per la sua unicità intrinseca: offre infatti la possibilità di caratterizzare dal punto di vista genetico una popolazione umana scomparsa a seguito del medesimo evento in una data ben documentata. Per tale ragione già alla fine degli anni '80 iniziarono i primi tentativi di recuperare ed analizzare il DNA estratto da reperti ossei ben conservati rinvenuti nella Casa dei Casti Amanti e da scheletri trovati nella casa di Caio Giulio Polibio [De Franciscis A. 1988]. Fu però nel 1998 che si ottennero i primi risultati genetici sul DNA estratto da femori ben conservati di alcuni scheletri della Casa di Polibio [Cipollaro M. 1998]. In questo lavoro è stata effettuata la determinazione genetica del sesso attraverso l'amplificazione di un locus presente in copia singola sugli eterosomi: il locus dell'amelogenina sul cr. X ed Y. Il locus del gene per l'amelogenina, una proteina

coinvolta nello sviluppo dello smalto dentale, è presente sia sul cromosoma X che sul cromosoma Y in posizione telomerica. Sul cr.X, però, il gene presenta a livello del primo introne una delezione di 6bp, assente nel gene sul cr.Y. Pertanto, amplificando per PCR questa regione sul cromosoma X si ottengono frammenti di 106bp, mentre gli ampliconi dal cr. Y sono di 112bp. A seguito dell'elettroforesi su gel di acrilammide dei prodotti di PCR amplificati da queste regioni corse insieme ad un marcatore di lunghezza, Cipollaro *et al.* identificarono il sesso di tre reperti in base al numero di bande visualizzate su gel: due bande (106bp e 112bp) -> sesso maschile; una banda (106bp) -> sesso femminile.

Questo metodo di determinazione genetica del sesso, però, può portare a falsi risultati perché frequentemente avviene l'amplificazione parziale di uno solo dei due alleli; inoltre, esiste una rara delezione del gene per l'amelogenina sul cr.Y che comporta l'assenza dell'amplicone di 112bp. Pertanto, la mancata amplificazione dell'allele di 112bp non può essere discriminante della presenza del cromosoma Y. Per fare diagnosi di genere è necessario mettere in evidenza sequenze che sono di pertinenza esclusiva del cromosoma Y.

1.5.3.4 Il DNA mitocondriale (mtDNA)

Nonostante l'avvento dell'analisi degli *SNPs* nucleari stia iniziando a migliorare le capacità di ottenere informazioni dal genoma nucleare endogeno di campioni antichi, questi molto frequentemente presentano uno stato di degradazione così spinto da rendere l'analisi del loro DNA nucleare molto improbabile. In tali casi, ed in generale nella prassi dello studio dell'adDNA, si cerca di ricavare informazioni in primo luogo dall'analisi del DNA mitocondriale, che sebbene sia meno informativa del DNA nucleare conferisce comunque un certo grado di informazione [Caramelli 2009; Butler 2009]. Nel caso di campioni antichi con DNA altamente degradato, la probabilità di ottenere risultati dall'analisi del DNA mitocondriale endogeno del campione è più alta rispetto a quella di ottenerli dal nucleare per le proprietà intrinseche al DNA mitocondriale stesso [Caramelli 2009; Butler 2009]. Una delle caratteristiche principali che rende più probabile il recupero di informazioni dall'mtDNA di campioni degradati è il più alto numero di copie per cellula con cui il DNA mitocondriale è presente. Mentre nel caso del DNA nucleare, infatti, ci sono solo due copie di esso per cellula (una materna e una paterna), il mtDNA può arrivare a contare centinaia o migliaia di copie per cellula aumentando la probabilità di riuscire a ricavare informazioni da qualcuna o da almeno una di esse [Butler 2009].

1.5.3.5 I mitocondri e le caratteristiche del mtDNA umano

I mitocondri sono organuli citoplasmatici delle cellule eucariotiche altamente specializzati per la funzione della respirazione cellulare: producono mediante la fosforilazione ossidativa grandi quantità di Adenosina Trifosfato (ATP) la molecola di scambio energetico che sostiene tutte le funzioni biochimiche delle cellule. Dal punto di vista strutturale sono rivestiti da due membrane separate dallo spazio intermembrana: la membrana mitocondriale esterna e la membrana mitocondriale interna che estendendosi nella matrice mitocondriale forma le creste mitocondriali (figura 14).

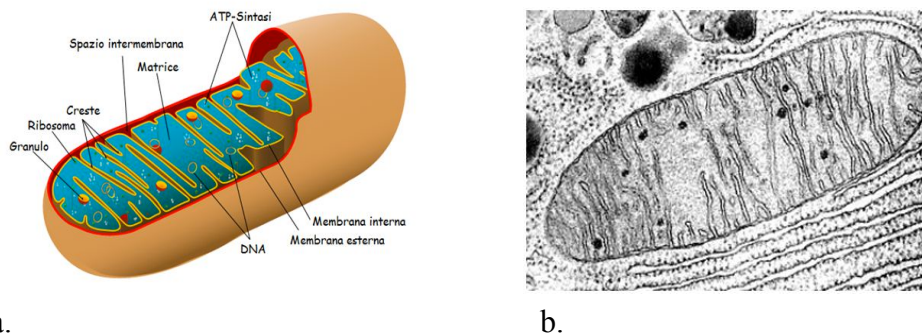


Figura 14: a. Rappresentazione schematica di un mitocondrio. b. Foto di un mitocondrio al microscopio elettronico a trasmissione (TEM).

Alcune caratteristiche dei mitocondri, quali le dimensioni di pochi micron, il rivestimento a doppia membrana, un genoma di tipo procariotico, ribosomi propri, capacità di moltiplicarsi per scissione binaria, sono alla base della teoria più accreditata della loro origine, l'ipotesi dell'endosimbionte. Secondo questa teoria, oltre un miliardo di anni fa, i mitocondri come procarioti liberi avrebbero instaurato un rapporto simbiotico di mutuo beneficio con organismi primitivi anaerobici. Con il tempo, questi batteri endosimbionti avrebbero perso o spostato materiale genetico nel nucleo dell'ospitante per la codifica di tutto ciò che non era più necessario o superfluo; nel contempo l'eucariote primitivo avrebbe avuto il vantaggio di trarre energia dalla fosforilazione ossidativa, portando all'evoluzione delle attuali cellule eucariotiche.

In una cellula umana, il 99.75% del genoma è rappresentato dal DNA nucleare, il restante 0.25% è contenuto nei mitocondri ed indicato come DNA mitocondriale (mtDNA) avente caratteristiche strutturali e funzionali diverse e indipendenti dal genoma nucleare [Nesheva 2014]. Il genoma mitocondriale si presenta in forma di una molecola di DNA circolare a doppio filamento superavvolta, non associata ad istoni o a proteine simili; i geni sono privi di introni e distanziati tra loro da nessuno o da pochi nucleotidi non codificanti. La replicazione dell'mtDNA è

semiconservativa, avviene per opera di DNA polimerasi specifiche del mitocondrio e non comprende meccanismi di correzione di bozze (*proof-reading*). Nei mitocondri delle cellule umane, possono essere presenti diverse copie di mtDNA (da 1 a 15 per mitocondrio); ogni cellula, poi, in base alle proprie necessità energetiche può contenere fino a centinaia di mitocondri per cellula, come nel caso della cellula uovo. In media è stato stimato che la maggioranza delle cellule umane ha 500 copie di mtDNA [Sato e Kuroiwa 1991]. Da ciò deriva l'elevata probabilità di riuscire ad analizzare questo tipo di DNA anche in campioni altamente degradati.

Relativamente alla sequenza, il genoma mitocondriale umano è costituito all'incirca da 16.569 bp, sebbene il numero di nucleotidi può variare da un individuo all'altro a causa di inserzioni o delezioni di una o più basi [Butler 2011]. L' mt DNA possiede 37 geni di cui 13 codificano proteine coinvolte nel processo della fosforilazione ossidativa o comunque nella respirazione cellulare e si trovano su entrambe le eliche; 2 geni codificano per RNA ribosomiali (rRNA) ed i restanti 22 per RNA transfer (tRNA) (figura 15).

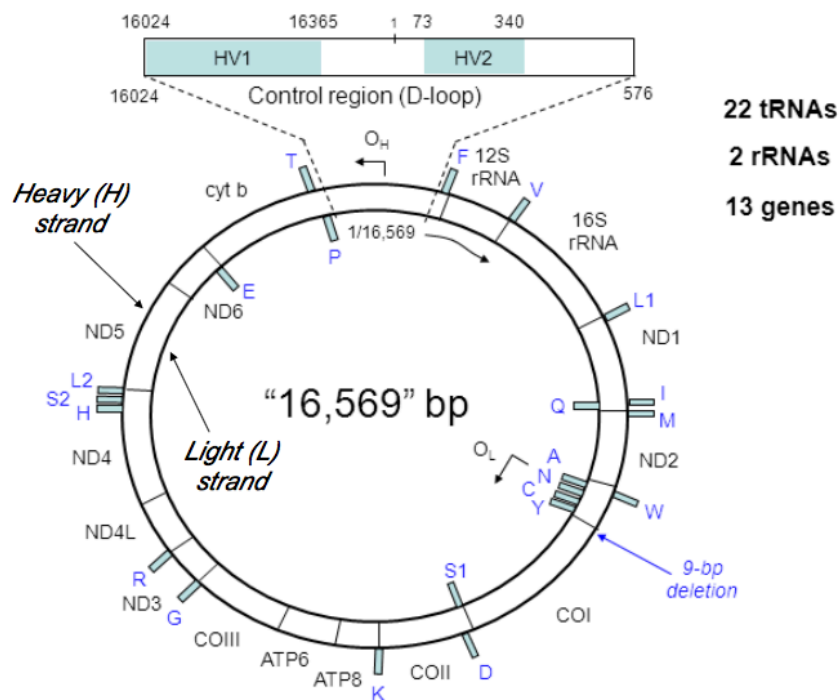


Figura 15: Rappresentazione schematica del genoma mitocondriale umano [Butler 2011].

Nel genoma mitocondriale possiamo poi riconoscere una distribuzione asimmetrica dei nucleotidi nelle due eliche complementari, le quali sono indicate come *Heavy (H)* e *Light (L)*, in base alla loro separazione per centrifugazione

all'equilibrio su gradiente di densità in CsCl. Nello specifico, il filamento *H* contiene un maggior numero di nucleotidi guanidinici che hanno peso molecolare relativo più alto rispetto agli altri residui nucleotidici.

C'è poi una regione di "controllo" di 1122bp che contiene l'origine di replicazione (O_H) del filamento *Heavy* e non contiene sequenze codificanti, pertanto è spesso indicata come "*non-coding*" region o *D-loop* (*displacement loop*) ove ha inizio anche la replicazione del mtDNA. Proprio perché la regione di controllo non codifica per funzioni necessarie alla vita della cellula, mostra una maggiore variabilità nella sequenza nucleotidica tra gli individui della specie. In particolare, la *D-loop* contiene al suo interno tre regioni che mostrano il maggior grado di polimorfismo: *hypervariable region I (HVI)*, *hypervariable region II (HV2)* ed *hypervariable region III (HV3)*, quest'ultima non mostrata in figura. Saltuariamente, in ambito forense, viene esaminata la terza regione ipervariabile (*HV3*) di 137bp fra 438-574 bp. Questo sito polimorfico aggiuntivo, in alcuni casi, può aiutare a distinguere campioni indistinguibili nell'*HVI* e 2 [Lutz R. et al. 2000].

Il sistema di numerazione delle posizioni nucleotidiche nel mtDNA considera come nucleotide 1 quello sul filamento *L* nel sito di restrizione per *MboI*, all'interno della regione di controllo, come definito nel lavoro originario che descrisse la sequenza del mtDNA umano (Anderson et al. 1981). La posizione 1 non è l'origine di replicazione.

1.5.3.6 Ereditarietà e polimorfismi dell' mtDNA

Il DNA mitocondriale nella specie umana ha trasmissione uniparentale materna: in una famiglia è trasmesso dalla madre a tutti i figli indipendentemente dal loro genere. Questo è dovuto agli eventi che avvengono all'atto della fecondazione nello zigote (2N) neoformatosi. I mitocondri dello spermatozoo umano sono localizzati nel così chiamato tratto intermedio della coda (figura16). Al momento della fecondazione solo la testa dello spermatozoo (contenente il nucleo, N, ma non i mitocondri) entra nella cellula uovo. Pertanto, la cellula uovo fecondata contiene soltanto mitocondri materni, che sono trasmessi alla progenie [Li R. 2015]. Occasionalmente, i mitocondri dello spermatozoo riescono ad entrare nella cellula uovo, ma questi vengono subito ubiquitinati per essere inviati ai sistemi di degradazione proteasomi e lisosomi nella cellula uovo fecondata [Sutovsky P., 2000; Li R. 2015].

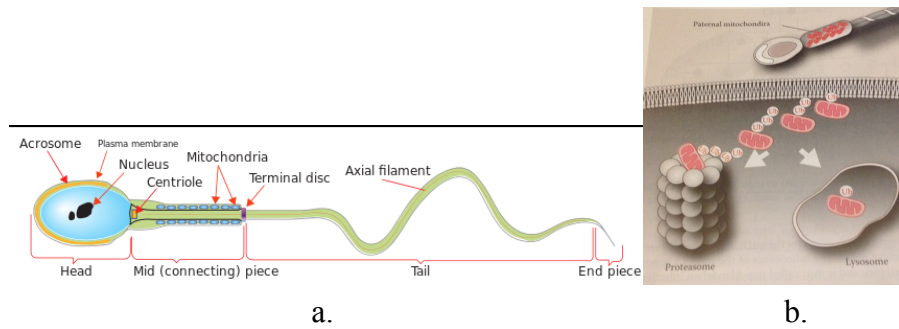


Figura 16: a. Rappresentazione schematica di uno spermatozoo umano. b. Degradazione dei mitocondri paterni eventualmente entrati nella cellula uovo fecondata [Li R. 2015].

Grazie a questo tipo di trasmissione, con il DNA mitocondriale è possibile costruire alberi filogenetici che seguono la discendenza matrilineare. Alla base dell'albero è posta un'antenata comune che ha trasmesso il proprio mtDNA praticamente inalterato, a meno di eventuali mutazioni. Secondo la teoria "dell'Eva Mitocondriale Africana" sembra che l'antenata comune a tutti i profili di mtDNA attualmente presenti sia una donna vissuta in Africa circa 200.000 anni fa [Vigilant et al., 1991]. Relativamente al tasso di variabilità del mtDNA, si stima che questo sia superiore a quello dei geni nucleari in singola copia per almeno un ordine di grandezza (10-20 volte più veloce in media) [Wallace et al, 1987]. In particolare, i polimorfismi che possono interessare il mtDNA sono fondamentalmente polimorfismi di sequenza per sostituzione di una o più basi e polimorfismi di inserzione/delezione di una o più basi. Le sostituzioni di basi possono essere transizioni purina - purina ($A \leftrightarrow G$) e pirimidina-pirimidina ($C \leftrightarrow T$) e sono quelle più frequenti; oppure transversioni purina-pirimidina e viceversa. Le sostituzioni nucleotidiche nel mtDNA originano, generalmente, da errori commessi dalla DNA polimerasi durante la fase di replicazione e che a causa della mancanza di attività *proof-reading*, rimangono fissati nel genoma mitocondriale. Prova ne è la maggior frequenza di questo tipo di polimorfismi nel mtDNA rispetto al DNA nucleare, ove la polimerasi è dotata di attività *proof-reading* [Russell P., 2006].

Le inserzioni o delezioni di una o più basi riscontrate nell' mtDNA, rendono conto delle differenze in lunghezza della sequenza del DNA mitocondriale tra un individuo e l'altro, come precedentemente accennato. Bisogna precisare, inoltre, che nelle regioni *HVs*, i siti polimorfici non sono distribuiti uniformemente, bensì sono localizzati in *clusters* chiamati "punti caldi". Dato l'alto numero di siti polimorfici, generalmente, l'analisi di *HV1* ed *HV2* rende possibile identificare l'appartenenza ad una data linea familiare materna.

Una regione caratterizzata da inserzioni e delezioni è ad esempio rappresentata dalla regione di controllo della subunità II della citocromo c ossidasi, analizzata negli studi popolazionistici perchè presenta una delezione di 9bp (indicata in

fig.15) nei nativi americani, negli abitanti delle isole del Pacifico e nelle zone del sud-est asiatico. La frequenza di questa delezione, unita ad altri marcatori è stata infatti utilizzata per investigare la colonizzazione delle isole del Pacifico [Hagelberg E., 1993].

1.5.3.7 Human mtDNA Reference Sequence(s) e Applicazioni dell'analisi del mtDNA

Il genoma mitocondriale umano fu sequenziato per la prima volta nel laboratorio di Fred Sanger all'Università di Cambridge, England, e la sequenza resa nota nel 1981 [Anderson et al. 1981]. Questa sequenza fu indicata come *Cambridge reference sequence (CRS)* in quanto sequenza di riferimento alla quale nuove sequenze di mtDNA potevano essere confrontate. La *CRS* era la sequenza di mtDNA derivata principalmente da un singolo individuo di origine europea. Successivamente, tecnologie di sequenziamento più accurate consentirono di ripsequenziare la *CRS* e di rettificare errori del precedente sequenziamento. La ***revised Cambridge Reference Sequence (rCRS)*** pubblicata nel 1999, rappresenta attualmente la sequenza standard di riferimento per i confronti di sequenze di mtDNA umane ed è disponibile in GenBank (NC_012920.1).

L'analisi dell'mtDNA viene ampiamente utilizzata in diverse discipline per scopi molto diversi tra loro. In ambito medico, sono state riscontrate strette correlazioni fra la presenza di mutazioni nell'mtDNA e l'insorgenza di patologie rare e multisistemiche in cui sono alterate le proteine della catena respiratoria [Wallace 1994].

Nel campo degli studi evolutivisti, i polimorfismi dell'mtDNA sono utilizzati per determinare le connessioni filogenetiche tra le specie. Un esempio di questo tipo è la scoperta che l'uomo di Neanderthal non è un antenato diretto dell'uomo moderno [Krings et al. 1997; Caramelli et al. 2003]. Dal confronto, infatti, della sequenza del mtDNA dei Neanderthal con quella dei Sapiens attuali e con quella dei primi Sapiens europei coevi dei Neanderthal (Cro-Magnon), è stato osservato che le sequenze di Neanderthal cadono al di fuori della variabilità genetica degli umani moderni e dei Cro-Magnon, pertanto, può essere escluso che i Neandertaliani siano i nostri diretti antenati.

In genetica di popolazione, gli Antropologi molecolari studiano le differenze nelle sequenze dell'mtDNA di vari gruppi di popolazioni per esaminare le loro origini e i fenomeni migratori che li hanno caratterizzati negli anni [Vernesi et al. 2004].

Un'applicazione interessante dell'analisi dell' mtDNA riguarda l'accertamento di relazioni di parentela per linea materna. Fra queste, un esempio storico è rappresentato dal caso della famiglia Romanov [Gill et al. 1994].

Anche in Genetica Forense può essere utilizzato l' mtDNA. Sebbene non utile ai fini dell'identificazione personale, l' mtDNA può rappresentare l'unico DNA ancora analizzabile in campioni altamente degradati o che per propria natura contengano una scarsa quantità di DNA: denti, capelli, ossa come le coste e le ossa lunghe (femori ed omeri). L' mtDNA non subendo ricombinazione e passando di generazione in generazione invariato dalla madre ai figli/e, è in grado di discriminare tutti gli individui appartenenti alla medesima linea parentale materna. Questo dato comunque è in grado di indirizzare le indagini ed è utilizzabile in tribunale [Faerman et al. 1997].

1.5.3.8 Aplogruppi umani dell' mtDNA

Con il termine aplotipo si indica la combinazione degli alleli in loci associati su un cromosoma o su un segmento di esso, tra i quali cioè non avvengono fenomeni di ricombinazione. Per quanto riguarda il DNA nucleare: sugli autosomi e sulle regioni pseudoautosomiche degli eterosomi, la presenza di loci associati può essere giustificata dalla vicinanza fisica tra i loci considerati e dall'assenza tra loro di *hot-spot* di ricombinazione.

Invece, gli alleli del genoma mitocondriale (mtDNA) sono sempre associati a formare aplotipi, così come gli alleli della regione non ricombinante del cromosoma Y (NRY). Come ricordato in precedenza, queste due porzioni genomiche non ricombinano, essendo ereditate con modalità uniparentale: materna la prima, paterna la seconda. Aplotipi differenti sono generati a partire da un aplotipo ancestrale per effetto di variazioni nucleotidiche ai singoli loci.

Per aplogruppo s'intende, invece, un insieme di aplotipi accomunati da un set di loci polimorfici recanti le medesime varianti alleliche ed ereditati da un progenitore comune. Dato un DNA mitocondriale umano, questo rappresenta sempre un aplotipo e secondo la teoria dell'Eva mitocondriale, tutte le linee mitocondriali osservate nelle popolazioni umane attuali discendono dall' mtDNA di una donna vissuta in Africa circa 150.000 anni fa [Cann et al 1987]. L'Eva mitocondriale, in altri termini, sarebbe il più recente antenato mitocondriale di tutti gli esseri umani viventi. La linea mitocondriale ancestrale si è evoluta dando origine a numerosi motivi mitocondriali, a tal punto che è possibile stabilire una forte connessione tra aplogruppi mitocondriali ed aree geografiche di origine [Wallace D.C. 1999]. Gli aplogruppi umani mitocondriali vengono indicati da una lettera maiuscola seguita da un numero identificante il sottogruppo [Richards et al 1998]. Il principale aplogruppo africano è indicato dalla lettera L e da esso si originano tutti gli aplogruppi di serie L. Di questi, L1 è considerato l'aplogruppo ancestrale perché localizzato alla radice dell'albero filogenetico, oltre ad essere il più diffuso in Africa; L3, invece, sembra essere il progenitore di almeno la metà

delle linee di mtDNA in Europa, Asia e Nativi Americani [E.C. Røyrvik 2016]. In particolare L3 da'origine ai macroaplogruppi M ed N (figura 17). M racchiude la prima migrazione umana fuori dall'Africa che ha raggiunto l'Asia circa 60 000 - 70 000 anni fa [Wallace D.C. 1999]. L'altro grande aplogruppo N, potrebbe rappresentare un'altra grande migrazione al di fuori del continente Africano, stavolta verso nord [Torrioni et al 1993].

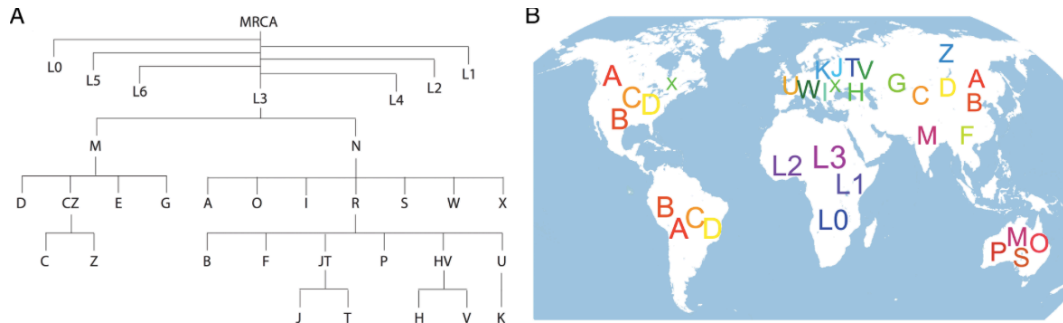


Figura 17: A. albero filogenetico degli aplogruppi umani mitocondriali. B. distribuzione geografica degli aplogruppi (mtDNA)MRCA = (mtDNA) Most Recent Common Ancestor). [E.C. Røyrvik 2016].

2. SCOPO DEL LAVORO

In tutta l'area archeologica di Pompei, negli strati di lapilli inferiori e nella parte superiore dei depositi piroclastici, sono stati recuperati finora, i resti scheletrici di migliaia di vittime dell'eruzione del 79 d.C. Per le modalità della morte e le condizioni caldo-umide dell'ambiente “*di sepoltura*”, per lungo tempo si è ritenuta altamente improbabile la presenza di DNA endogeno nei resti scheletrici recuperati dagli scavi. Sebbene negli anni '90 del secolo scorso e agli inizi degli anni 2000 siano stati effettuati degli studi mediante metodiche tradizionali su alcuni resti scheletrici ben conservati di Pompeiani [Cipollaro M 1998; Di Bernardo 2009]; non era stata effettuata prima del presente lavoro di Dottorato, l'analisi del genoma mitocondriale e nucleare completi da campioni ossei altamente degradati delle vittime del 79 d.C.

Nel presente lavoro, in particolare, abbiamo analizzato campioni ossei altamente degradati prelevati in fase di restauro dei rispettivi calchi in gesso, allestiti nel sito archeologico di Pompei. Si tratta di sedici reperti scheletrici in calco e due reperti dentali, recuperati in diverse aree dell'antica Pompei: *Casa del Bracciale d'Oro*; *Casa del Criptoportico*; *Macellum*; *Villa dei Misteri*; *Terme Stabiane*; *Terme del Sarno*; *Granai del foro*; così come indicato nella mappa a fine paragrafo (figura 18).

I campioni ossei analizzati sono rappresentati da due reperti dentali, unici campioni ben conservati, e da sedici frammenti ossei non integri, di dimensioni ridotte, talvolta irrisorie, ed in condizioni di conservazione estremamente precarie nonché frammisti al gesso del calco, come descritto in dettaglio nella sezione Materiali e Metodi.

Da un tale materiale di partenza, il mio lavoro di Dottorato aveva lo scopo di valorizzare la conoscenza di questi reperti attraverso analisi genetico-molecolari volte a verificare la presenza di DNA endogeno ancora analizzabile e caratterizzare dal punto di vista genetico questa piccola popolazione. Allo scopo, abbiamo effettuato sia analisi sul genoma mitocondriale attraverso le quali abbiamo potuto accertare anche eventuali relazioni di parentela per linea materna, sia analisi sul DNA nucleare al fine di ottenere molteplici informazioni sugli individui in esame, quali la determinazione genetica del sesso e l'origine geografica.

Nel presente lavoro di Dottorato le analisi genetico-molecolari dei campioni sono state organizzate come segue.

1. Analisi del DNA mitocondriale (*capture-on-beads method* ed *NGS -MiSeq sequencing*) al fine di indagare eventuali relazioni di parentela per linea materna;
2. Analisi del DNA nucleare:
 - determinazione genetica del sesso (*DNA shotgun sequencing*);
 - Cattura e sequenziamento di 1.240.000 *SNPs* da ciascun campione, al fine di collocare anche in termini geografici gli individui esaminati (*Solution-phase Hybrid Selection* ed *NGS- HiSeq sequencing*);
3. Analisi isotopiche, per identificare eventuali migrazioni effettuate dai soggetti in esame durante la propria vita o dai loro predecessori (rapporti degli isotopi dello Stronzio e del Piombo).
4. Analisi radiometriche a conferma della datazione dei reperti esaminati e quindi dell'autenticità dei dati ottenuti (^{14}C).

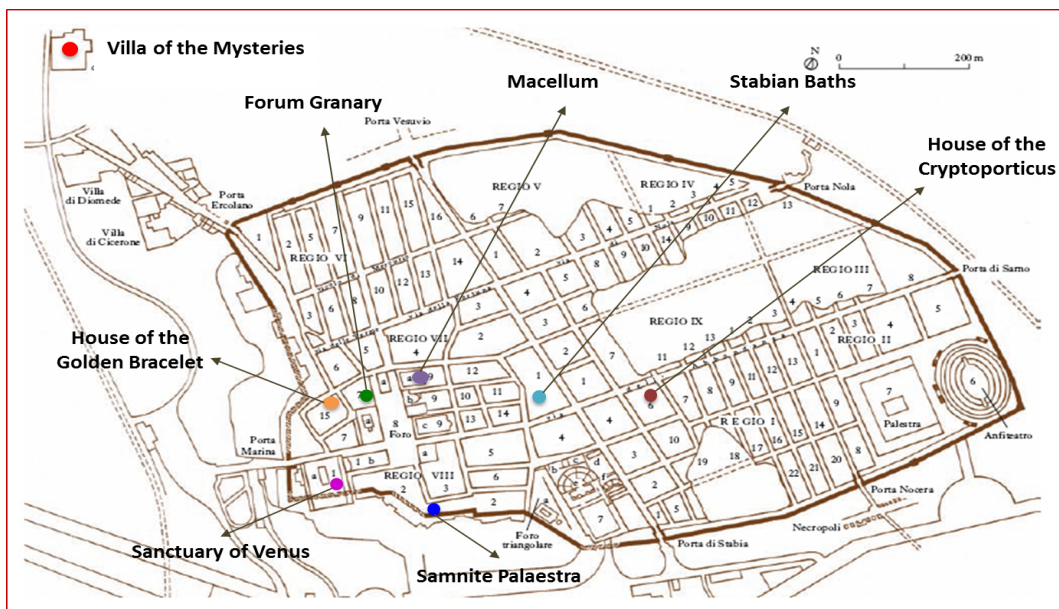


Figura 18: Mappa dell'area archeologica di Pompei con indicazione dei siti in cui sono stati recuperati i reperti ossei analizzati nel presente lavoro.

3. MATERIALI E METODI

3.1 Campionamenti

Nel presente lavoro di Dottorato sono stati analizzati sedici frammenti ossei e due campioni dentali prelevati durante il restauro dei calchi di alcune vittime dell'eruzione del 79 d.C. a Pompei.

Si tratta di resti umani recuperati nel corso delle operazioni di scavo in case gentilizie di note famiglie del patriziato romano ed in centri nevralgici dell'area sud-occidentale dell'antica città, quali le terme ed il *macellum*.

Le *Domus* in questione sono la Casa del Bracciale d'oro e quella del Criptoportico.

Quattro dei resti scheletrici esaminati, appartenevano ad individui ritrovati nello stesso ambiente della Casa del Bracciale d'oro, imponente abitazione signorile dell'*Insula Occidentalis*.



a.



b.

Figura 19: a. Il bracciale che ha dato il nome all'omonima casa. B. Locale interno della casa con i calchi dei quattro individui recuperati ed analizzati nel presente lavoro.

Localizzata nel Vico del Farmacista, la Casa del Bracciale d'Oro fu riportata alla luce per la prima volta già a metà '700, in epoca borbonica, ma le effettive opere di scavo cominciarono intorno al 1974, quando si iniziarono a definire i confini dei vari complessi abitativi di quella zona della città fino ad allora indicati col nome generico di "Case di Fabio Rufo". Caratteristica distintiva di questa *Domus* è la disposizione degli ambienti su tre livelli che rendono l'intera struttura proiettata in verticale piuttosto che in orizzontale. Alla base è presente un ampio giardino con vista sul mare e sul golfo di Napoli. La dimora prende il nome dal ritrovamento di un bracciale d'oro dall'eccezionale peso di seicentodieci grammi: nel corso dello scavo nel settore di servizio, sono stati rinvenuti quattro individui,

di cui due adulti e due infanti che si erano rifugiati nel sottoscala dove trovarono la morte [Ciardiello 2012]. Al braccio di uno dei due adulti fu rinvenuta la preziosa armilla (figura 19). Di questi quattro individui è stato possibile realizzare il calco, da cui in fase di restauro sono stati prelevati frammenti ossei irrimediabilmente misti a gesso e indicati rispettivamente come campioni C-21, C-34, C-23, C-35 (figure 19 e 20).

Gli Archeologi, poi, esaminando i reperti ossei avevano supposto si trattasse di un nucleo familiare rimasto unito negli ultimi istanti in un ambiente interno della casa per cercare rifugio dalle polveri vulcaniche.

Ancora nella Casa del Bracciale d'Oro, sono stati recuperati gli scheletri di due individui adulti i cui calchi dopo il restauro hanno mostrato che erano rimasti stretti tra loro come in un abbraccio negli ultimi istanti della loro vita (figura 22). Gli Archeologi avevano supposto fossero due sorelle rimaste abbracciate nel tentativo di proteggersi. Le analisi genetiche, pertanto, rappresentano l'effettivo mezzo per verificare queste ipotesi. Ai fini del presente lavoro, sono stati prelevati un frammento osseo (C-32) da un soggetto, ed un dente, campione C-33, dall'altro soggetto (figura 22).

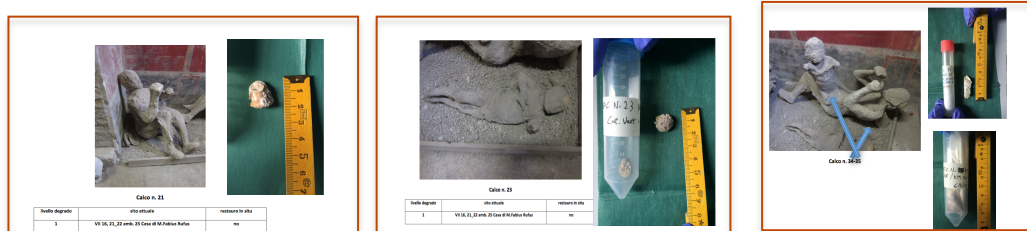


Figura 20. Da sinistra, calco e frammento osseo campione C-21; calco e frammento osseo campione C-23; calchi e frammenti ossei prelevati per l'analisi rispettivamente indicati C-34 e C-35.



Figura 21. Calchi prima del restauro delle due vittime recuperate nella Casa del Bracciale d'Oro con i relativi campioni prelevati ai fini delle analisi: a sinistra C-32; a destra C-33.



Figura 22. Calchi delle presunte sorelle della Casa del Bracciale d’Oro dopo il restauro.

Altri campioni recuperati durante il restauro di calchi nella Casa del Bracciale d’Oro sono indicati di seguito (figura 23).



Figura 23: Calchi e campioni (C-6; C-7; C-8) recuperati nella Casa del Bracciale d’Oro.

Anche nella Casa del Criptoportico sono stati recuperati frammenti scheletrici per le analisi eseguite nel presente lavoro. Posta su via dell’Abbondanza con ingresso secondario su vicolo del Menandro, la casa deve la sua moderna denominazione alla presenza, al di sotto dell’ampio giardino quadrangolare, di un lussuoso criptoportico (corridoio sotterraneo voltato) con grandi finestre, sul quale si aprivano una stanza di soggiorno (oecus) e quattro ambienti termali coperti da volte originariamente decorate in fine stucco (figura 24). Nel 1914, durante lo scavo del giardino, furono ritrovati i resti di vittime dell’eruzione, forse gli abitanti della casa: di questi, furono realizzati i calchi e da uno di essi in fase di restauro è stato prelevato il materiale scheletrico per l’analisi genetica, campione C-11 (figura 24).



Figura 24. A sinistra fotografia del Criptoportico. A destra il calco del pompeiano con relativo campione prelevato per l'analisi, C-11.

Altri resti scheletrici analizzati nel presente lavoro sono stati recuperati nei seguenti siti: le terme Stabiane, le terme del Sarno, il *macellum*, i granai del foro, il santuario di Venere e la Villa dei Misteri.

A Pompei si trovavano ben quattro impianti termali pubblici entro le mura della città ed uno poco fuori le mura, oltre ad impianti privati anche di grande dimensione, come quello della Casa del Criptoportico. A partire dalla fine del IV sec.a.C. Pompei si dotò, dapprima delle cosiddette Terme Stabiane, l'edificio termale più antico della città, situato in una delle zone maggiormente frequentate, tra via dell'Abbondanza e via di Stabia (figura 25). Successivamente furono realizzate le Terme del Sarno, mentre le Terme Centrali erano ancora in fase di costruzione al tempo dell'eruzione del Vesuvio nel 79 d.C.. Da calchi in restauro di individui recuperati nelle terme Stabiane e del Sarno sono stati prelevati i campioni d'osso C-77 e C-66, C-67, rispettivamente (figura 26).

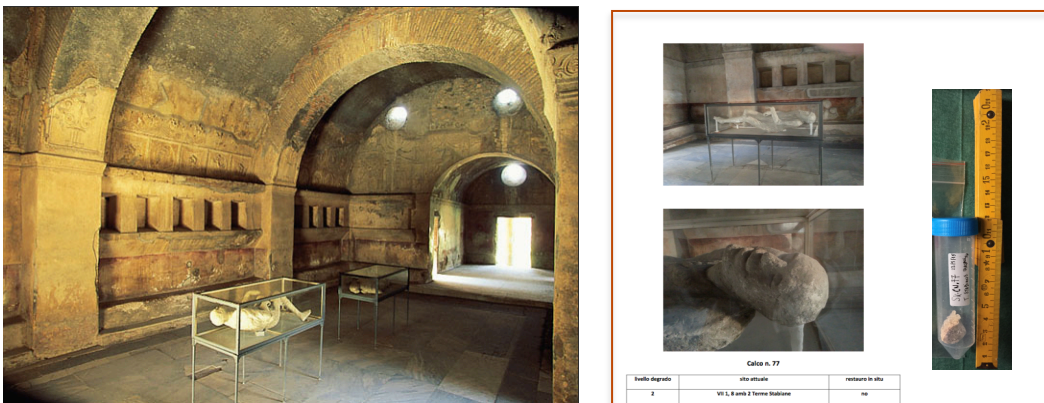


Figura 25. Terme Stabiane e calco da cui è stato prelevato il frammento osseo, campione C-77, per le analisi genetiche.

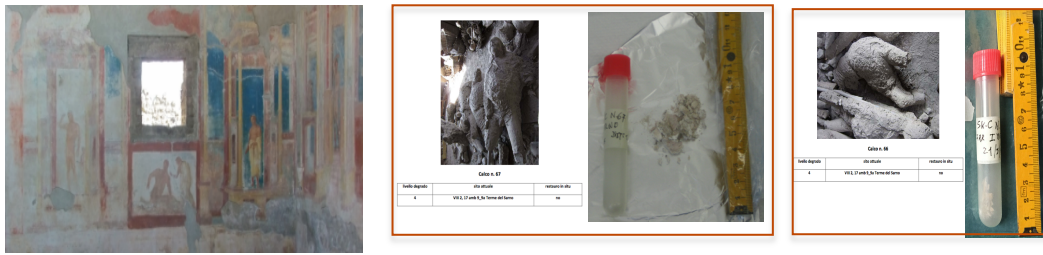


Figura 26. Dipinti nella Palestra delle terme del Sarno (a sinistra). Calchi e frammenti ossei, campioni C-66, C-67, utilizzati per le analisi.

Il *macellum*, rappresentava poi, il principale mercato della città e fu edificato con successive ristrutturazioni nel II sec. a.C.. Le basi dinanzi al portico d'ingresso reggevano statue onorarie di cittadini illustri (figura 27). In figura sono riportati i calchi delle vittime recuperate nel *macellum* e da cui è stato prelevato il frammento osseo utilizzato per le analisi, campioni C-30 e C-31.



Figura 27. Ingresso del *Macellum* di Pompei (a sinistra). Calchi e campioni C-30 e C31.

Qualche centinaio di metri fuori dalle mura nord dell'antica città è collocata la più nota villa suburbana di epoca romana di Pompei, la Villa dei Misteri. Si tratta di un edificio a carattere rustico-residenziale riportato alla luce a partire dagli inizi del XX secolo, è uno degli edifici più visitati soprattutto per la serie di affreschi del triclinio, raffiguranti riti misterici ben conservati (qui di fianco un particolare), da cui la struttura prende il nome (figura 28). In figura 29 è riportato il calco della vittima della grande eruzione recuperata nella Villa dei Misteri da cui è stato prelevato il reperto dentale C-5.

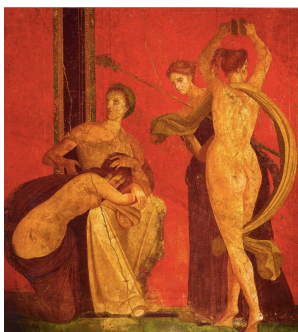




Figura 28. A sinistra sala del triclinio della villa dei misteri decorata dal grande dipinto a tema dionisiaco che ha dato il nome alla residenza. A destra particolare del grande affresco della sala.

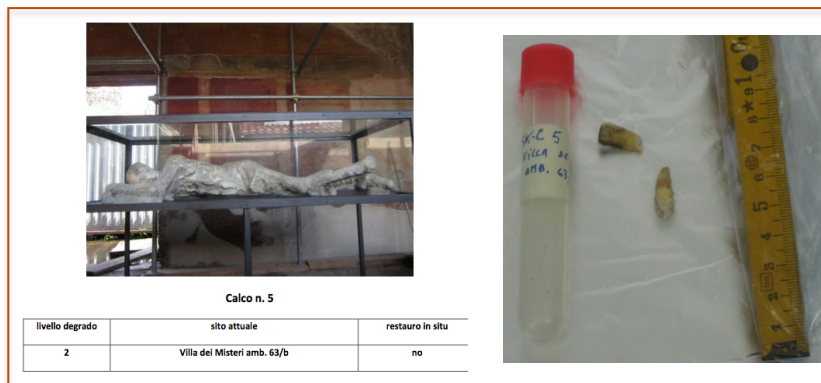


Figura 29. Calco e campione C-5, prelevato per le analisi.

Altri campioni prelevati per le analisi con i relativi calchi sono indicati in figura 30.



Figura 30. A sinistra calco di una vittima recuperata nei granai del foro e relativo campione, C-27; A destra calco e campione, C-86, dal santuario di Venere.

Quadro sinottico dei campionamenti:

Individuo	Sito di recupero	Campione
Adulto	Casa del bracciale d'oro	C-21
Adulto	Casa del bracciale d'oro	C-34
Infante	Casa del bracciale d'oro	C-35
Infante	Casa del bracciale d'oro	C-23
Adulto	Casa del bracciale d'oro	C-32
Adulto	Casa del bracciale d'oro	C-33 (dente)
Adulto	Casa del bracciale d'oro	C-8
Adulto	Casa del bracciale d'oro	C-6
Adulto	Casa del Bracciale d'oro	C-7
Adulto	Santuario	C-86
Adulto	Granai del foro	C-27
Adulto	Casa del criptoportico	C-11
Adulto	Terme stabiane	C-77
Adulto	Terme del sarno	C-66
Adulto	Terme del sarno	C-67
Adulto	Macellum	C-30
Adulto	Macellum	C-31
Adulto	Villa dei Misteri	C-5 (dente)

Tabella 2: descrizione dei campionamenti con indicazione del sito di recupero dei calchi in restauro considerati nel presente lavoro.

3.2 Ambiente di lavoro e prevenzione delle contaminazioni

Le analisi genetiche sono state condotte presso due laboratori strutturati per lo studio del DNA antico: il laboratorio di Antropologia Molecolare dell'Università degli Studi di Firenze, ove è stato condotto il lavoro sperimentale per l'analisi del DNA mitocondriale; ed il laboratorio del Prof. David Reich al dipartimento di Genetica dell'Università di Harvard, in Boston Massachusetts (USA), per le analisi sul genoma nucleare. Il problema principale del lavoro sperimentale sul DNA antico è la contaminazione da DNA moderno, per questo motivo, entrambi i laboratori coinvolti nelle analisi sono organizzati secondo specifici criteri e operano seguendo una serie di norme atte a ridurre al minimo questo rischio. Il laboratorio di Antropologia Molecolare dove ho svolto l'attività sperimentale del mio lavoro di Dottorato, è dotato di due aree fisicamente separate: l'area di **pre-amplificazione (Pre)** e l'area di **post-amplificazione (Post)** con accesso indipendente e preposte a fasi diverse del lavoro sperimentale. Nell'area di pre-amplificazione (*Pre*) si effettua tutta la fase sperimentale che precede l'amplificazione del DNA: pulizia e polverizzazione dei reperti ossei, estrazione del DNA e preparazione della librerie genomiche. Questa zona del laboratorio dispone di una *dressing-room* in cui l'operatore indossa tuta, camice sterile, mascherina, cuffia per capelli, calzari e guanti sterili al fine di controllare in modo molto stringente le contaminazioni. In quest'area tutte le superfici di lavoro vengono usualmente lavate con ipoclorito di sodio e irraggiate con radiazioni UV a 254 nm, prima e dopo l'utilizzo. Anche gli strumenti, altri materiali e la vetreria, sono sterilizzati con ipoclorito di sodio o acido acetico al 30% (per materiali in metallo) e raggi UV254 nm. Particolare attenzione è adottata per tutti i reagenti utilizzati nel *Pre* che vengono aliquotati in sterilità, evitando così l'uso ripetuto e prolungato da una singola *stock*. Durante le varie fasi di lavoro, poi, vengono analizzati i controlli negativi, contenenti solo la miscela dei reagenti senza il DNA, sia per l'estrazione che per la libreria genomica, così da poter monitorare costantemente la presenza/assenza di contaminanti. Per evitare contaminazioni, inoltre, è richiesto che reagenti e strumenti dell'area *Post*, non vengano mai introdotti nell'area *Pre* e gli operatori, entrati nel *Post* non possono più entrare nella zona *Pre* nell'arco della stessa giornata. Nell'area di post-amplificazione (*Post*), invece, vengono eseguite tutte le fasi di lavoro su DNA amplificato e quindi meno esposto a contaminazioni. In particolare, vengono eseguite le analisi elettroforetiche e di clonaggio adottate nella metodologia classica e le fasi finali del flusso di lavoro previste dal protocollo *Next Generation Sequencing* (Illumina): *Real Time PCR*, arricchimento e cattura del DNA mitocondriale.

3.3 Pulizia e polverizzazione dei campioni

La pulizia dei reperti e la loro polverizzazione sono state effettuate nel laboratorio *Pre*, sotto cappa aspirante, lavata con ipoclorito di sodio e sterilizzata con raggi UV prima e dopo l'utilizzo.

Una volta giunti in laboratorio, i campioni sono stati accuratamente puliti per eliminare dalla superficie eventuali residui di terra, polvere e porzioni di gesso del calco, nonché materiale biologico di natura epiteliale, depresso dagli operatori che hanno manipolato i campioni durante le fasi di scavo e recupero degli stessi. Dato che è stato sperimentalmente dimostrato che il DNA contaminante può penetrare la superficie delle ossa fino a uno spessore di 1- 2 mm [Bouwman et al., 2006], questo spessore superficiale viene rimosso utilizzando un trapano odontoiatrico munito di punte metalliche precedentemente lavate con acido acetico al 30%, acqua bidistillata e sterilizzate con gli UV per 45 minuti. Una volta rimosso lo strato superficiale, ogni superficie dei campioni è stata sottoposta a irraggiamento con UV per 45 minuti all'interno di un *crosslinker* (*Biolink DNA crosslink*). A questo punto ho proseguito con la fase di polverizzazione: attraverso l'utilizzo dello stesso trapano odontoiatrico con microfresse sterili, ho praticato un foro nel campione per ottenere una quantità di polvere sufficiente alla successiva fase di estrazione del DNA. Il protocollo di estrazione utilizzato, in particolare, richiede 50 mg di polvere d'osso al fine di un'estrazione efficiente. A causa dell'elevato stato di degradazione dei frammenti ossei, da alcuni campioni non è stato possibile ricavare la quantità minima di polvere richiesta. Pertanto, per i campioni: C8; C7; C27; C66; C31 non è stato possibile proseguire alla successiva fase di estrazione. Per gli altri 13 campioni la polvere d'osso prodotta è stata trasferita all'interno di *ependorf Safe lock* e conservata in freezer a -20° C al fine di preservare il materiale.

3.4 Estrazione del DNA da polvere d'osso

Il processo mediante il quale si ottiene l'estrazione del DNA dalle cellule, si basa sulla rottura delle membrane cellulari mediante solubilizzazione progressiva delle strutture cellulari che lo contengono, denaturazione delle proteine ed inattivazione di enzimi che potrebbero degradarlo. Si effettua un trattamento con soluzioni contenenti tensioattivi, proteasi ed agenti chelanti i sali di calcio che costituiscono la struttura ossea. L'estrazione è stata effettuata nella sezione di *Pre-amplificazione* del laboratorio, sotto la stessa cappa a flusso laminare utilizzata per pulizia e polverizzazione, indossando guanti chirurgici e adottando tutte le norme già elencate per assicurare l'assoluta sterilità del processo. Al fine di realizzare un controllo altamente stringente delle contaminazioni, in questa fase

delicata del lavoro sperimentale, è stato incluso un controllo negativo (k – estrazione), che permetterà poi di stabilire se gli estratti sono contaminati o meno.

Il protocollo di estrazione seguito è quello pubblicato da Dabney et al. [Dabney et al. 2013] proposto come evoluzione del precedente protocollo pubblicato da Rohland e Hofreiter [Rohland e Hofreiter 2007]. A differenza di quest'ultimo, il protocollo proposto da Dabney necessita di una minore quantità di polvere d'osso (circa 50 mg contro i 150-250 mg richiesti dal protocollo precedente). Pertanto, il protocollo di Dabney, oltre a massimizzare l'estrazione di frammenti corti, congeniale per effettuare un'analisi sul DNA antico altamente degradato, necessita anche di un minor materiale di partenza, caratteristica fondamentale nel caso dei campioni ossei in esame di ridotte dimensioni e notevolmente frammentati. Il protocollo *Dabney* prevede l'utilizzo della guanidina idrocloride (GuHCl), sodio acetato, isopropanolo e tween 20, colonnine MinElute (*Qiagen*) dotate di una membrana di silice idratata, materiale che presenta elevata affinità per gli acidi nucleici a pH acidi. Inoltre è un protocollo veloce che si articola in due giorni.

Per realizzare l'estrazione è necessario allestire tre reagenti:

- 1) **EXTRACTION BUFFER**, con il quale vengono lisate le cellule per consentire la fuoriuscita del contenuto cellulare ed è composto da EDTA 0.5 M a pH 8, che chela gli ioni Ca^{2+} e Mg^{2+} favorendo la decalcificazione della componente inorganica della matrice ossea, e l'inattivazione di enzimi litici, le DNasi in particolare, che utilizzano lo ione Mg^{2+} come cofattore. Proteinasi K, 10 mg/ml: una endopeptidasi; Acqua di grado HPLC; Tween20: un detergente che solubilizza i lipidi di membrana favorendo la lisi cellulare.
- 2) **BINDING BUFFER**, che favorisce il legame del DNA alla membrana di silice della colonnina MinElute (*Qiagen*) ed è costituito da Guanidina idrocloride (5M concentrazione finale), un sale caotropico che consente l'adsorbimento del DNA alla membrana di silice idratata della colonnina; Acqua di grado HPLC; Isopropanolo (40% concentrazione finale) che favorisce la precipitazione del DNA rimuovendo la sfera di solvatazione delle molecole d'acqua attorno ai gruppi fosfato della doppia elica; Tween 20 (0.05% conc. finale).

Entrambi i buffer sono stati sterilizzati sotto UV per 45' prima dell'utilizzo.

- 3) **TET BUFFER**, che favorisce l'eluizione del DNA promuovendone il distacco dalla membrana di silice della MinElute. E' composto da EDTA 0.5 M a pH 8 chelante gli ioni magnesio e quindi capace di inattivare le

nucleasi proteggendo il DNA dalla degradazione; Acqua di grado HPLC; Tween 20; Tris-HCl 1M, pH8. Quest'ultima è una soluzione tampone che in ambiente alcalino favorisce la deprotonazione dei gruppi fosfato della doppia elica e dei gruppi imminici delle basi azotate; il DNA perde, quindi, affinità per la resina e diviene nuovamente solubile in acqua.

Per ciascun campione è stato aggiunto 1ml di Extraction Buffer alla eppendorf *Safe Lock* contenente la polvere d'osso (circa 50 mg). Le eppendorf sono state poi centrifugate per risospingere la polvere, sigillate con *parafilm* e incubate *over-night* a 37°C su un rotore, in modo tale da favorire la digestione della polvere e il rilascio del DNA.

Il giorno successivo le eppendorf riprese dal rotore sono state centrifugate a 15300 rpm per 2', in maniera tale da separare la polvere che si depositerà sul fondo (*pellet*) dal DNA che rimarrà in soluzione, rappresentando il surnatante. I pellet dei campioni sono stati conservati in freezer a -20°C per eventuali estrazioni future, mentre i surnatanti trasferiti in falcon sterili da 15 ml, ove per ciascun campione e per il k-estrazione, sono presenti 10 ml di *Binding buffer* e 400µl di Sodio acetato 3M a pH 5,2, un sale che annullando le cariche negative dei gruppi fosfato, favorisce l'interazione tra il DNA e la membrana di silice della colonnina MinElute. Parallelamente è stato preparato un apparato (figura 31) costituito dal Reservoir, rimosso da una Zymo-Spyn V column (Zymo Research), precedentemente lavato in ipoclorito di sodio, acqua e sterilizzato con UV per 45', ed è stato forzato nell'apertura di una colonnina MinElute (Quiagen) privata del *collection tube*; l'apparato è stato poi, inserito all'interno di una falcon sterile da 50 ml e, prima dell'utilizzo, sottoposto a centrifugazione a 1500 rpm per 4' (accelerazione 5 e decelerazione 9) al fine di valutarne la stabilità.



Figura 31: Schema dell'apparato Reservoir-Zymo/MinElute.

La soluzione composta da *Binding buffer*, Sodio acetato e surnatante è stata trasferita nel Reservoir. La falcon contenente l'apparato, chiusa col tappo a vite è stata centrifugata prima a 1500 rpm per 4' e poi, ruotata di 90° e centrifugata a 1500 rpm per 2'. Quindi, rimosso dalla falcon, l'apparato è stato posto nel *collection tube* della MinElute, e dopo aver rimosso il Reservoir, la colonnina, con il DNA legato alla membrana di silice, è stata centrifugata a secco a 6000 rpm per 1'. Quindi sono stati effettuati due lavaggi consecutivi della colonnina, aggiungendo ogni volta 750 µl di Buffer PE (Qiagen) e centrifugando a 6000 rpm per 30'' in modo da eliminare eventuali impurità. Eliminato il liquido percolato, la colonnina è stata ruotata di 180° e centrifugata due volte a secco alla velocità massima (15300 rpm) per 1'. Quindi la MinElute è stata posta in una nuova eppendorf siliconata da 1,5 ml per le fasi di eluizione. Vengono aggiunti 50 µl di *TET buffer* sulla membrana della MinElute e trascorsi 5' la colonnina viene centrifugata alla massima velocità per 30''. Questo passaggio è stato ripetuto nuovamente così da ottenere per ciascun campione un volume finale eluito e quindi di estratto pari a 100 µl. Gli estratti sono stati conservati in freezer a -20°C prima di essere utilizzati nelle fasi successive.

3.5 Preparazione delle librerie genomiche *half-UDG*

Per la preparazione delle librerie genomiche è stato seguito il protocollo pubblicato da Meyer e Kircher [Meyer e Kircher 2010]. Inoltre, per ridurre il tasso di Uracili nelle molecole d'interesse in modo da diminuire gli errori nella chiamata delle basi in fase di sequenziamento ed al tempo stesso lasciare tali modificazioni nucleotidiche distintive delle molecole antiche, è stato effettuato nelle fasi iniziali della preparazione delle librerie, un trattamento parziale con l'enzima Uracil DNA Glicosidasi (UDG), in maniera tale da consentire la persistenza di questo tipo di modificazione soltanto alle estremità delle molecole. Infatti, è stato sperimentalmente dimostrato che a seguito di questo trattamento le molecole endogene dei campioni antichi mostrano un tasso di errore C->T superiore al 3% confinato alle estremità [N. Rohland, E. Harney 2014].

Tutte le fasi in cui si articola questo processo sono state eseguite nella zona di *Pre-amplificazione* del laboratorio, seguendo tutte le norme descritte per ridurre al minimo le contaminazioni e mantenere l'assoluta sterilità. Il protocollo si articola nelle fasi di *Blunt-end repair e trattamento half-UDG*, *Adapter Ligation*, *Adapter Fill-in*, *Indexing PCR* indicate in figura 32 e descritte di seguito.

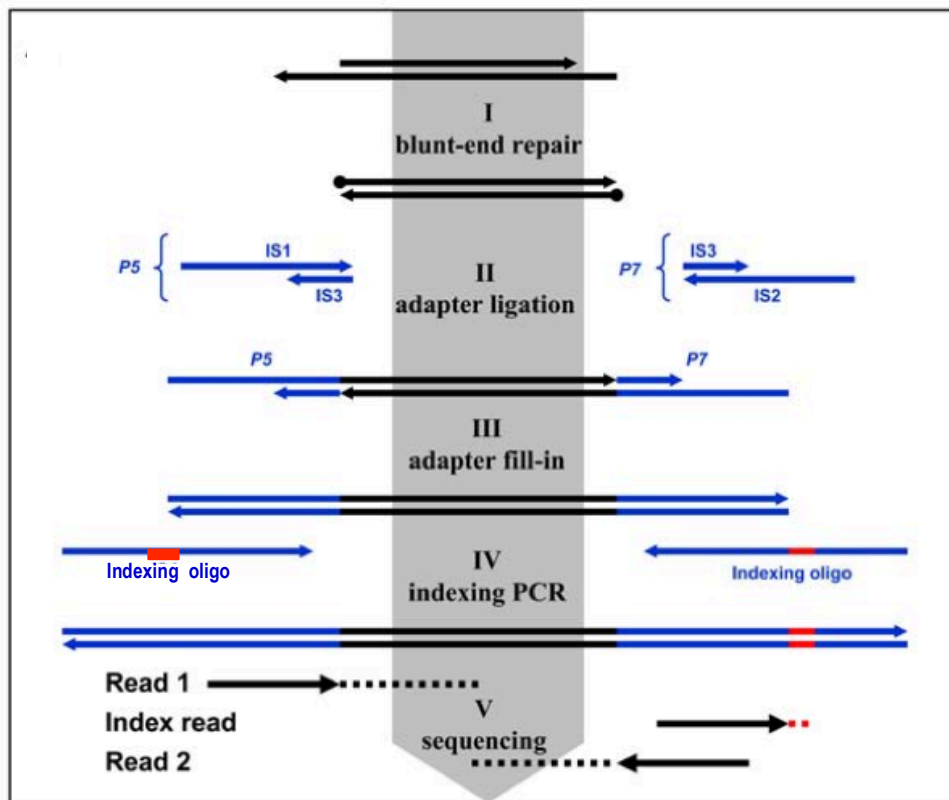


Figura 32: Schema delle fasi della preparazione delle librerie genomiche [Meyer e Kircher 2010]. Schema modificato da Francesca Amitrano.

3.5.1 Blunt-end repair e trattamento half-UDG

Questa prima fase di preparazione delle librerie genomiche, prevede il riparo delle estremità sporgenti a singolo filamento alle estremità 5' e 3' dei frammenti estratti, in modo da generare estremità piatte funzionali al successivo legame degli adattatori. La reazione di riparo avviene ad opera della T4 polimerasi e della T4 polinucleotide chinasi (PNK): la PNK catalizza il legame di un gruppo fosfato all'estremità 5'-OH del frammento, mentre la T4 polimerasi guida l'estensione delle estremità a singolo filamento fino a generare estremità piatte (*Blunt-ends*). La mix di reazione è preparata in tubini per PCR da 0,2 ml, ha un volume finale di 52,2 μ l e per ciascun campione e controllo è composta da: 6 μ l di buffer Tango 10x; 0,24 μ l di dNTP mix 25 mM; 0,6 μ l di ATP 100 mM; 3,6 μ l di USER enzyme mix (1 U/ μ l, NEB) contenente l'enzima Uracil-DNA-glicosidasi; 20 μ l di DNA estratto o acqua nel caso del controllo negativo; acqua fino a volume.

La reazione è stata incubata in un termociclatore impostato a 37° C per 30'. Durante questo periodo la maggior parte delle citosine deamminate vengono escisse dalla UDG e i siti abasici tagliati via dalla Endonucleasi VIII. Trascorso questo tempo, ad ogni tubino sono stati aggiunti 3,6 µl di inibitore dell'enzima UDG, Uracil Glicosidasi Inibitore (UGI) (2 U/µl, UDG inhibitor, NEB) e quindi reincubati nuovamente a 37° C per 30'. A questo punto, alle miscele di reazione sono stati aggiunti 3 µl di T4 PNK (10 U/µl) ed 1,2 µl di T4 Polymerase (5 U/ µl). I risultanti 60µl finali in ciascun tubino, sono stati incubati in termociclatore con il seguente profilo termico:

- 25°C per 15',
- 12°C per 5'.

I prodotti ottenuti vengono purificati utilizzando il MinElute *Purification Kit* (Qiagen). Ad ogni campione, caricato nell'apposita colonnina MinElute, si aggiungono 300 µl di PB *buffer*, si centrifuga a 14000 rpm per 1 minuto e si scarta il liquido percolato. Si aggiungono poi 700 µl di PE *buffer* alle colonnine che di nuovo vengono centrifugate a 14000 rpm per 1 minuto. Una volta scartato il liquido percolato, le colonnine vengono ruotate di 180° prima di essere centrifugate nuovamente 1 minuto a 14000 rpm, in questo modo si eliminano i residui di PE in tutti i punti del filtro della colonnina. Le colonnine vengono, quindi, trasferite in nuove *eppendorf* siliconate sterili da 1,5 ml e si aggiungono 18 µl di Tris-HCl 10mM nel centro del filtro. Dopo 5 minuti gli apparati sono centrifugati per 1 minuto a 14000 rpm. L'eluato, 18 µl di DNA *Blunt-end repaired*, è pronto per passare alla fase successiva.

3.5.2 Adapter ligation

In questa fase si legano alle estremità della molecola di DNA gli adattatori Solexa Illumina specifici per il sequenziamento (P5 e P7). Gli adattatori Solexa Illumina sono preparati a partire da tre oligonucleotidi :

IS1_adapter_P5.R (5'-AATGATACGGCGACCACCGA),
IS2_adapter_P7.F (5'- CAAGCAGAAGACGGCATACTGA) e
IS3_adapter_P5+P7 (3'-AGATCGGAAGAGC).

Il protocollo di preparazione della *Solexa Adapter Mix* prevede la costituzione di due adattatori (P5 e P7) solo parzialmente a doppio filamento, essendo l'oligonucleotide IS3_adapter_P5+P7 complementare solo a una porzione degli altri due oligonucleotidi P5 e P7 che danno il nome agli adattatori. Questa particolare conformazione, con una sola estremità piatta, garantisce che gli adattatori si leghino alle molecole di DNA del campione nell'orientamento

corretto. Gli adattatori P5 e P7 si preparano separatamente e poi vengono riuniti in un' unica mix: *Solexa Adapter Mix*. Per la reazione di *Adapter ligation* è stata preparata una miscela di reazione avente volume finale di 21 μl e composta, per singolo campione, da 20 μl di *Quick Ligase Buffer (2X)* ed 1 μl di *Solexa Adapter Mix* (10000 nM). Questa miscela è stata aggiunta ai 18 μl di DNA *Blunt-end* o di controllo negativo, e ad ogni tubino è stato aggiunto 1 μl di *Quick ligase* (stock conc. 5 U/ μl , final conc. 0,125U/ μl), l'enzima che catalizza il legame degli adattatori. La reazione è stata incubata per 20' a temperatura ambiente. Avvenuta la reazione, si è proceduto alla purificazione utilizzando il *MinElute Purification Kit* eluendo in 16 μl di TET.

3.5.3 Adapter Fill-in

Con questo passaggio si estendono le porzioni degli adattatori ancora a singolo filamento e si colmano eventuali interruzioni del legame fosfodiesterico tra l'adattatore ed il frammento di DNA ad opera della *Bst* polimerasi. Per questa fase è stata allestita una mix che per singolo campione e controllo è composta da: 4 μl di *Isothermal buffer (10X)* (stock conc. 10X, final conc 1X), 2 μl di dNTP 2,5 mM (stock conc. 2500 nM, final conc 125 nM), 2 μl di *Bst polymerase 2.0* (stock conc. 8 U/ μl , final conc. 0,4 U/ μl), 16 μl di Dna stampo eluito dall'*Adapter ligation*, oppure acqua nel caso del controllo negativo, ed 1 μl di acqua per portare ad un volume finale di 25 μl per tubino.

La reazione è stata incubata nel termociclatore impostato con il seguente profilo termico:

- 37°C per 20 minuti
- 80°C per 20 minuti.

Al termine della reazione i prodotti della *Fill-in* sono stati direttamente trasferiti in *eppendorf* siliconate sterili senza purificazione, con *MinElute Purification Kit* e congelati.

3.5.4 Indexing PCR e qPCR

In questa fase le librerie vengono indicizzate scegliendo combinazioni di indici diverse per ciascun campione. Questo consente di sequenziare insieme (*multiplex*) librerie appartenenti a campioni diversi e successivamente, durante l'analisi delle sequenze, discriminare un campione dall'altro in base alla coppia di indici assegnata.

Ogni indice è costituito da sequenze lunghe 6 o 7 bp ed è inserito all'interno di un *indexing-primer* che ha da un lato la sequenza complementare a quella degli adattatori (P5 e P7) e dall'altro una sequenza utilizzabile per le successive fasi di amplificazione e quantificazione.

Per la reazione di indicizzazione è stata preparata una miscela avente un volume finale di 100 µl, che, per singolo campione contiene: 10 µl di *Buffer* 10X (stock conc. 10X, final conc, 1X); 1 µl di dNTP mix 25 mM (stock conc. 25 mM, final conc. 0,25 mM); 1,5 µl di BSA (stock conc. 20X, conc finale 0,3X), 1 µl di *Pfu Turbo Polymerase* (conc stock 2,5 U/ µl, final conc. 0,025 U/ µl). Ciascun campione è stato preparato in quattro aliquote (*split*) in tubini da 0,2 ml sterili. In ciascun tubino sono stati aggiunti 82 µl della miscela di reazione preparata, 10 µl della libreria eluita dalla *Fill-in* e 4 µl di ciascun indice (stock conc. 10 mM, final conc. 0,2 mM) per un volume finale di 100 µl. I tubini sono stati poi incubati nel termociclatore, impostando il seguente profilo termico:

-*Step* di attivazione: 95°C per 2’;

Per 20 cicli:

-*Step* di denaturazione: 95°C per 30’’;

-*Step* di *annealing*: 58 °C per 30’’;

-*Step* di amplificazione: 72°C per 1’;

Al termine dei 20 cicli, *Step* finale di estensione: 72°C per 10’.

Conclusa la reazione, le librerie indicizzate sono state purificate con *MinElute Purification Kit (Quiagen)*, come descritto in precedenza. Durante la purificazione le quattro aliquote di un medesimo campione vengono riunite in una sola colonnina MinElute aggiungendo 500 µl di PB *buffer* per ogni aliquota. Si eluisce in un volume di 50 µl di TET.

A questo punto l’efficienza della *Indexing PCR* è stata valutata mediante *Real-Time PCR* (RT-PCR). La RT-PCR è una PCR quantitativa (qPCR) che permette, grazie all’utilizzo di un fluoroforo che si intercala tra le basi del DNA, di monitorare la fase esponenziale della reazione di PCR e quantificare il contenuto di DNA presente in ogni libreria. La miscela di reazione è preparata usando il *kit SsoFast EvaGreen (Bio-Rad)*, che utilizza il *SYBER green* come fluoroforo, e i *primers Solexa Primer Fwd IS5* e *Solexa Primer Rvs IS6*, complementari alla sequenza degli *indexing-primers* esterna a quella degli indici. Le quantità di reagenti per ciascun campione (volume finale 10 µl) sono: 5 µl di *supermix EvaGreen 2x*, 0.5 µl di ciascun primer e 3 µl di acqua. Nella *plate*, insieme ai campioni vengono caricati in doppio degli standard a concentrazione nota e la libreria è stata diluita 1:100 e 1:1000 per farla ricadere nella retta standard, mentre i controlli negativi sono stati caricati tal quale. La *plate* caricata con librerie e standard, viene posta nell’apposito strumento (MiniOpticon *Real-Time PCR System*, Biorad) e si utilizza il seguente profilo termico:

- Attivazione iniziale: 95°C per 15’;

Per 40 cicli:

- Denaturazione: 94°C per 10’’;
- *Annealing*: 53 °C per 30’’;
- Estensione: 72 °C per 30’’;

Al termine, curva di *Melting*.

Per ciascun campione e standard, è stata quindi ottenuta una curva di amplificazione, espressione dell'intensità di fluorescenza (proporzionale alla quantità di amplificato prodotto dalla reazione) in funzione del numero di cicli della PCR ($Y=fX$). Utilizzando i valori di concentrazione iniziale degli standard e ponendo il valore *Ct* (*Threshold cycle*) uguale al numero di cicli a cui si registra un aumento della fluorescenza (determinato dalla quantità di campione presente all'inizio della reazione di amplificazione) in funzione del log ([DNA] iniziale), si costruisce una retta di taratura da cui si estrapola la concentrazione dei campioni in esame.

La RT-PCR fornisce, dunque come risultato, delle quantificazioni che sono state utilizzate per effettuare alcuni calcoli necessari per la successiva fase di arricchimento. In un foglio di calcolo *Excel* vengono riportati i valori ottenuti per tutte le librerie diluite e mediante formule preimpostate sono stati calcolati i relativi valori di concentrazione (copie/ μ l), la cui media è stata moltiplicata per il volume di libreria rimasto (volume totale della reazione di *indexing* a cui vengono sottratti i 2 μ l usati per la RT-PCR), ottenendo così il numero di copie totali di ciascuna libreria. Questo valore è necessario per stabilire il numero di aliquote in cui deve essere suddiviso il campione nella successiva fase di arricchimento ed il numero di cicli idonei affinché non venga superato il valore di 10^{13} molecole di DNA, che determinerebbe l'incorporazione di errori durante la fase di amplificazione e la formazione di prodotti aspecifici.

3.6. Arricchimento delle librerie mediante PCR

L'arricchimento delle librerie di DNA è stato effettuato nell'area *POST* del laboratorio, mediante PCR utilizzando il kit *Herculase* (*Agilent Technologies*). La miscela di reazione, preparata in un volume finale di 50 μ l contiene per singolo campione: 10 μ l di *buffer* 5x, 0,5 μ l di dNTP mix 25 mM, 2 μ l di ciascun primer IS5 e IS6 10 μ M, 0,5 μ l di *Herculase II Phusion DNA polymerase*, acqua sterile per portare a volume. In base al numero di aliquote stabilite per ciascun campione, la mix è stata aliquotata in tubini da 0,2 ml a cui sono stati aggiunti 5 μ l di libreria.

Il profilo termico utilizzato per la reazione è il seguente:

- Attivazione: 95°C per 2';
- Denaturazione: 95°C per 15''
- Annealing*: 60°C per 30'';
- Amplificazione: 68°C per 1';

Al termine, amplificazione: 68°C per 5'.

I prodotti della reazione sono stati purificati con il *MinElute Purification Kit* (*Qiagen*) e analizzati all'*Agilent Bioanalyzer 2100*, usando il kit DNA1000, al fine di stimare il profilo qualitativo e quantitativo dei campioni prima di procedere alle fasi successive: cattura del DNA mitocondriale presso il laboratorio di Antropologia molecolare a Firenze; sequenziamento *shotgun* a bassa copertura presso il dipartimento di Genetica dell'Università di Harvard.

3.7 Cattura del DNA mitocondriale

Il DNA mitocondriale delle librerie *half-UDG* di ciascun campione è stato catturato mediante l'applicazione del metodo proposto da Maricic e colleghi nel 2010 [Maricic et al. 2010]. Questo metodo prevede l'utilizzo di sonde di mtDNA umano moderno, rappresentate da due prodotti di PCR *long-range* parzialmente sovrapposti che coprono l'intero mtDNA. I prodotti delle PCR *long-range* vengono sottoposti ad ultrasonificazione per ottenere frammenti di lunghezza compresa tra 150 e 300 bp. Tali frammenti sono biotinilati su entrambe le estremità e denaturati ed i singoli filamenti ottenuti sono poi immobilizzati su biglie magnetiche rivestite da streptavidina. L'ibridazione delle sonde immobilizzate sulle biglie con le librerie *half-UDG*, viene realizzata in due notti di incubazione. Trascorso questo tempo le molecole delle librerie che non hanno ibridato alle sonde vengono lavate via, mentre quelle che hanno ibridato e quindi che sono rimaste catturate sulle biglie, sono eluite e quantificate mediante RT-PCR ed eventualmente arricchite prima di essere sequenziate con *Next Generation Sequencing (NGS)*. Lo schema del protocollo Maricic è riportato in figura 33.

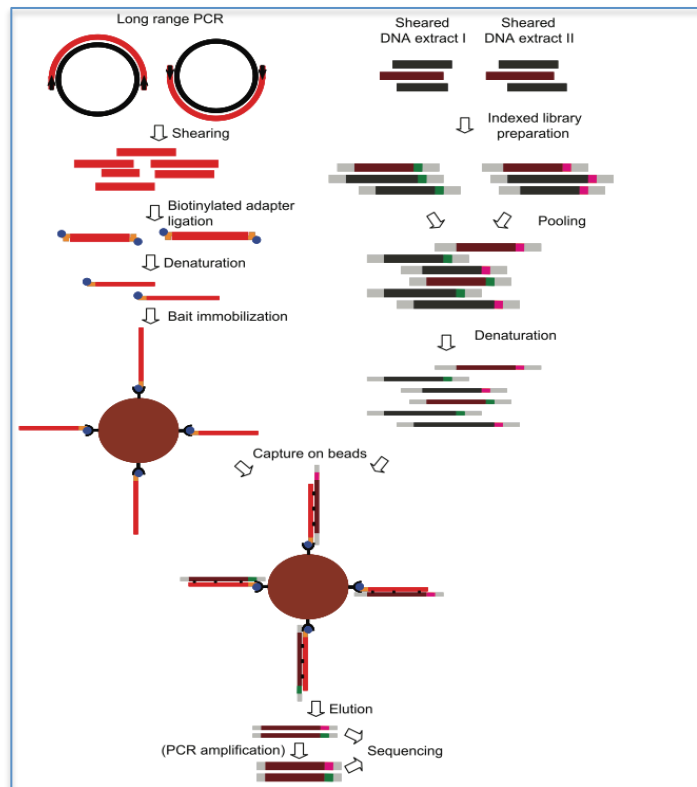


Figura 33: Schema riassuntivo del metodo di cattura del mtDNA proposto da Maricic *et al.* nel 2010 [Maricic *et al.* 2010].

3.7.1 Preparazione delle sonde

Le sonde per la cattura del mtDNA delle librerie, sono state preparate in laboratorio a partire da DNA umano moderno estratto da tampone buccale. L'estrazione è stata realizzata mediante il *QIAamp Investigator Kit* (Qiagen). Il tampone è stato posto in una *ependorf* da 2 ml a cui sono stati aggiunti 20 µl di proteinasi K, 400 µl di *buffer* ATL e successivamente 400 µl di *buffer* AL. Il tubino posto in un *thermomixer* è stato incubato a 70°C con agitazione 900 rpm per 10'. Quindi, sono stati aggiunti 200 µl di etanolo alla soluzione che è stata poi sottoposta a vortex per 15". Il lisato trasferito in una colonnina *QIAamp MinElute* (Qiagen) dotata di *collection tube*, è stato centrifugato a 6000 x g (8000 rpm) per 1'. Al termine, la colonnina trasferita in un nuovo *collection tube*, è stata caricata dapprima con 500 µl di *buffer* AW1 e poi con 700 µl di *buffer* AW2 centrifugando ogni volta a 6000 x g (8000 rpm) per 1'. Quindi, la colonnina è stata trasferita in un nuovo *collection tube* e sono stati aggiunti 700 µl di etanolo al 100%; dopo una prima centrifuga a 6000 x g (8000 rpm) per 1', è stata eseguita un'ulteriore centrifuga a secco ed alla velocità massima (20000 x g, 14.000 rpm) per 3' in modo da eliminare eventuali residui di etanolo dalla membrana.

A questo punto la colonnina è stata trasferita in una *ependorf* da 1,5 ml e lasciando il tappo aperto è stata incubata a temperatura ambiente per 10'. Trascorso questo tempo sono stati aggiunti al centro della membrana della colonnina 50 µl di *buffer ATE* al fine di eluire il DNA legato alla membrana di silice. Atteso 1' la colonnina è stata centrifugata alla velocità massima per 1'. L'estratto ottenuto è stato quantificato al *NanoDrop 1000 (Thermo Scientific)* e utilizzato per allestire due PCR *long-range* con il *Kit Expand Long Range System (Roche)*, in modo tale da ottenere due amplificati che sovrapponendosi parzialmente coprono l'intero genoma mitocondriale.

Il protocollo di amplificazione prevede la preparazione di una miscela di reazione avente un volume finale di 50 µl, che per singolo campione è composta da: 10 µl di *buffer Expand Long Range System (Roche) 5x*, 1,5 µl di DMSO al 100%, 1,5 µl di ciascun primer 10 µM, 2,5 µl di dNTP mix 10 µM, 3,7 U di *Enzyme Expand Long Range System (Roche)*, 1 µl di BSA 20 mg/ml, 30,3 µl di acqua e 1 µl di estratto. Seguendo le istruzioni del produttore è stato impostato il seguente profilo termico:

- 92°C per 2' (attivazione)

Per 40 cicli si ripetono i seguenti step termici:

-Denaturazione: 92°C per 10''

-*Annealing*: 68°C per 15''

-Estensione: 68°C per 9'

Al termine dei cicli:

-Estensione: 68°C per 7'

Per realizzare questa amplificazione sono utilizzati due coppie di *primers* diversi aventi le seguenti sequenze:

Prima coppia di *primers*:

Fwd: 5'-GGCTTTCTCAACTTTTAAAGGATA

Rvs: 5'-TGTCCTGATCCAACATCCA

Seconda coppia di *primers*:

Fwd: 5'-CCGTGCAAAGGTAGCATAATC

Rvs: 5'-TTACTTTTATTTGGAGTTGCACCA

I prodotti delle PCR *Long Range*, sono stati controllati mediante elettroforesi su gel d'agarosio e purificati con il *MinElute Purification kit* eluendo in 130 µl di *TE low Buffer*. I purificati sono stati trasferiti in tubini appositi per l'ultrasonicazione (*Covaris microtube*) e sottoposti a sonicazione con il sistema *Bioruptor* per ottenere frammenti di DNA di circa 300 bp, controllati nuovamente mediante elettroforesi su gel di agarosio. Come descritto in precedenza, il protocollo di

Marcic prevede l'utilizzo di biglie magnetiche ricoperte da streptavidina come supporto per le sonde, che pertanto, devono essere sottoposte a trattamenti in modo che possano aderire alla superficie delle biglie. Il primo processo consiste nel riparo delle estremità a singolo filamento dei frammenti sonicati (*Blunt-end Repair*). Pertanto, 76 μ l di frammenti sonicati sono stati incubati a temperatura ambiente per 30 minuti in una miscela di reazione composta da: *Blunting Buffer* 1x, una mix di dNTP 0,1 mM e *Blunting Enzyme mix*, in un volume finale di 100 μ l. Dopo la purificazione con *MinElute Purification Kit (Quiagen)*, le sonde vengono biotinilate mediante il legame a specifici adattatori biotinilati (*Adaper Bio-T/B*) ad opera dell'enzima *Quick Ligase*, in una reazione che avviene a temperatura ambiente per 15 minuti. Al termine della reazione i prodotti sono stati nuovamente purificati con il *MinElute Purification kit* e poiché è stato calcolato che per ogni ng di biglie riescono a legarsi al massimo 25 ng di DNA, i prodotti purificati sono stati quantificati al *Nanodrop* così da conoscere la concentrazione delle sonde e massimizzarne il legame alle biglie.

3.7.2 Legame delle sonde alle biglie magnetiche

Il legame delle sonde di mtDNA alle biglie magnetiche è mediato dall'interazione ad alta affinità tra biotina e streptavidina: le biglie magnetiche sono funzionalizzate con streptavidina, proteina in grado di legare con altissima affinità la biotina degli adattatori legati alle sonde nella fase precedente. Poiché la streptavidina è in grado di staccarsi più facilmente dalle biglie quando è legata alle molecole di DNA è necessario preparare le biglie rivestite dalle sonde il giorno stesso in cui si esegue l'esperimento di cattura. Da quanto scritto nel precedente paragrafo, si potrebbe utilizzare un volume di biglie pari a 5 μ l per legare 125 ng di sonde; il protocollo Maricic, però, riporta che l'efficienza combinata di legame tra sonde ed adattatori e tra questi e le biglie è al massimo del 30%, quindi è necessario disporre di 400-500 ng di sonde biotinate per poter saturare completamente i 5 μ l di biglie. Queste ultime sono state inizialmente sottoposte a due fasi di lavaggio: il primo con 200 μ l di *buffer BWT 1X*, (costituito da NaCl 2M, Tris-HCl 10 mM pH8, EDTA 1 mM pH8, Tween 20 allo 0,1% e H₂O); il secondo con 200 μ l di TET.

D'altro canto, le sonde biotinate il cui volume è compreso tra 5 e 50 μ l, sono state denaturate in un ugual volume di *buffer BWT 2X* in un termociclatore a 98°C per 1 minuto e subito poste in ghiaccio per evitare il riappaiamento tra i filamenti complementari. A questo punto le sonde biotinate e denaturate, vengono miscelate alle biglie mediante un'incubazione a temperatura ambiente per 20 minuti su di un rotore, che agitando i tubini, facilita l'interazione biotina-streptavidina e quindi il legame sonda-biglia. Al termine dei 20', attraverso l'uso

di un *rack* magnetico (*DynaMag Spin, Invitrogen*) si recuperano le biglie e si rimuove il surnatante contenente le sonde non legate. A questo punto le biglie sono state sottoposte a due lavaggi con 200 μ l di *Buffer BWT 1X* riscaldato a 50°C e un lavaggio finale con 50 μ l di TET. Quindi sono state conservate a 4°C. A questo punto le biglie sono rivestite di sonde atte alla cattura del DNA mitocondriale delle librerie.

3.7.3 Ibridazione tra le sonde e le librerie

Affinchè le sonde denaturate possano ibridare con il DNA mitocondriale delle librerie indicizzate, queste ultime devono essere rese a singolo filamento e i filamenti complementari non devono riappaiarsi tra loro. A questo scopo il protocollo Maricic prevede una fase preliminare nella quale le librerie sono incubate con *Blocking Oligo*, oligonucleotidi complementari agli adattatori *Solexa Illumina P5* e *P7*. Grazie alla presenza degli indici legati ai frammenti di DNA delle librerie nelle fasi precedenti, si possono aggiungere campioni diversi in una medesima reazione di cattura, in modo tale da utilizzare una quantità totale di DNA pari a 2000 ng equivalente al volume finale di 16,8 μ l richiesto per la preparazione della *Blocking-Oligo Hybridization Mixture*. Per calcolare il volume di ciascun campione (libreria) da aggiungere in una medesima reazione di cattura è stato considerato il valore di concentrazione ottenuto dal profilo *Agilent* nell'intervallo di lunghezze 150-400 bp. Passaggio, questo, fondamentale per un'efficiente cattura: i campioni inseriti nella stessa reazione devono essere fra loro equimolari. Quindi si prepara una *Master Mix* con un volume finale per singola cattura pari a 35,2 μ l e che contiene: 1 μ l di BO4.part1.R (stock conc. 100 μ M); 1 μ l di BO6P7.part2.R (stock conc. 100 μ M); 1 μ l di BO8.P5.part1.R (stock conc. 100 μ M); 1 μ l di BO11.sslib.part2.R (stock conc. 100 μ M); 5,2 μ l di *Agilent blocking agent*; 26 μ l di *Agilent hybridization buffer* (stock conc. 2X).

Il volume di *Hybridization Mix* ed i 16,8 μ l totali di librerie (volume di 52 μ l), posti in tubini di 0,2 ml, sono stati incubati nel termociclatore per 3 minuti a 95°C e poi 30 minuti a 37°C. Terminata questa fase, ho proceduto alla cattura vera e propria. Dopo aver rimosso il TET nel quale erano state risospese e conservate le biglie magnetiche rivestite dalle sonde, sono stati aggiunti i 52 μ l di *Hybridization Mix* più librerie. Dopo aver risospeso bene la miscela di reazione, questa è stata trasferita in *ependorf Safe Lock* da 0,5 ml sigillate con *Parafilm e Cap locks* in modo da evitare la fuoriuscita di liquido durante l'incubazione che è stata realizzata su un rotore a 12 rpm posto in stufa a 65°C per 48 ore.

3.7.4 Eluizione delle librerie catturate

Trascorso il periodo di incubazione, si effettuano una serie di fasi di lavaggio per eliminare la parte di libreria non ibridata. Con l'ausilio di un *rack* magnetico (*DynaMag Spin, Invitrogen*) le biglie con l'mtDNA legato, vengono separate dal resto della soluzione ed il surnatante scartato. Successivamente le biglie vengono lavate tre volte con 200 μ l di *Buffer BWT 1X*, eliminando di volta in volta il surnatante ed utilizzando il magnete. Le biglie, poi, sono state risospese e incubate per due volte in 200 μ l di *Buffer HWT*, composto da MgCl 2,5 mM, KCl 50 mM, Tris-HCl 15 mM a pH 8, 0,1% di Tween-20, preriscaldato a 60°C per 2 minuti. A questo punto le biglie legate al DNA *target* sono state risospese in 100 μ l di TET e trasferite in nuove *ependorf* da 1,5 ml per procedere all'eluizione della libreria catturata. Nella fase di eluizione sono stati utilizzati 50 μ l di una soluzione di NaOH 125 mM capace di rompere i legami a ponte d'idrogeno instauratisi tra i filamenti di sonda e di DNA *target*, rilasciando in soluzione l'mtDNA catturato. Avvalendosi del magnete si recuperano le biglie, mentre il surnatante contenente il DNA *target* viene trasferito in nuovi tubini. Questi passaggi sono stati ripetuti utilizzando altri 50 μ l di NaOH 125 mM al fine di massimizzare il recupero del DNA *target* eluito. Per evitare la degradazione del DNA, la soluzione di eluizione alcalina viene immediatamente neutralizzata aggiungendo una soluzione tampone preparata con 500 μ l di *PBI (kit MinElute, Qiagen)* e 3,8 μ l di acido acetico al 20%. Il *PBI* contiene un indicatore di pH che permette di verificare il valore di pH della soluzione in base al viraggio del colore dal giallo (soluzione tendente all'acidità) al blu/violaceo (soluzione tendente all'alcalinità). Il viraggio del colore verso il blu/violaceo è stato tamponato con l'aggiunta di ulteriore ac. acetico al 20%. Gli eluiti sono stati quindi purificati utilizzando il *MinElute Purification Kit (Qiagen)*, eluendo i catturati in 15 μ l di TET e conservandoli a -20°C.

3.7.5 Quantificazione e arricchimento del DNA catturato

Per verificare la resa della cattura è stata eseguita una *Real Time PCR* utilizzando i *primers Bridge IS5* e *IS6*, complementari alle sequenze esterne degli *indexing-primer*, con il medesimo kit e protocollo precedentemente descritti diluendo, però, i catturati 1:10 in quanto il processo di cattura riduce drasticamente la quantità di DNA. Anche in questo caso il risultato della qPCR è stato utilizzato per stabilire il numero di cicli da impostare nella successiva fase di arricchimento, che prevede l'utilizzo del *kit Herculase (Agilent Technologies)* seguendo il medesimo protocollo descritto in precedenza. Successivamente, le catture di tutte le librerie sono state riunite insieme a formare una miscela equimolare pari a 2 nM e sottoposte a sequenziamento ultramassivo su piattaforma *Illumina Mi-Seq*.

3.8. *The Next Generation Sequencing Technologies: piattaforme*

Illumina

Le più moderne e utilizzate tecnologie di sequenziamento dei genomi, indicate complessivamente come *Next Generation Sequencing (NGS)* sono sistemi di sequenziamento ad alta resa in grado di generare un altissimo numero di sequenze (Gigabasi) in poche ore ed a costi contenuti. Queste caratteristiche le rendono funzionali all'analisi di genomi completi o geni molto lunghi ed il loro perfezionamento sta rivoluzionando la ricerca in diverse aree della Biologia. La loro efficienza e versatilità le rende adatte anche all'analisi del DNA antico, ove hanno quasi completamente sostituito il metodo di Sanger. Con *NGS* è possibile sequenziare un campione di DNA altamente degradato limitando tutte le problematiche del caso: frammentazione, degradazione e basso numero di copie, con enormi vantaggi rispetto alla metodologia classica. Nello studio del DNA antico le piattaforme *NGS* maggiormente utilizzate sono state, inizialmente la 454/Roche FLX e successivamente la *Illumina Genome Analyser*, attualmente in uso.

La prima, descritta nel 2005 [Margulies et al. 2005] a causa delle procedure più complesse e dei costi più elevati, a parità di *output* di sequenziamento, è stata rapidamente sostituita da Illumina.

Nonostante l'elevata sensibilità e l'enorme numero di dati che è in grado di generare, il sistema di rilevamento delle piattaforme *Illumina* rappresentato da una camera *CCD (Charge-coupled device)*, non è capace di rilevare il segnale relativo ad una singola molecola di DNA, ma lo individua solo se è generato da milioni di molecole. Per questo nel flusso di lavoro è fondamentale la costruzione di librerie genomiche di DNA target e la loro amplificazione, come descritto nei paragrafi precedenti.

Le fasi caratteristiche del *work flow NGS Illumina* sono fondamentalmente tre:

1. **Preparazione di librerie genomiche**, che, come descritto, prevede alle estremità dei frammenti di DNA l'inserzione di adattatori universali, brevi oligonucleotidi di sequenza nota sui quali sono disegnati i *primers* necessari alle successive fasi di amplificazione e sequenziamento.

2. **Amplificazione delle librerie**, in cui i frammenti vengono amplificati utilizzando *primers* complementari agli adattatori. In particolare, l'amplificazione avviene su un supporto solido, detto *flow cell*, costituito da una *slide* otticamente trasparente formata da 8 canali (*lanes*), su cui sono legati gli oligonucleotidi di ancoraggio complementari agli adattatori delle librerie (figura 34). In questo modo le molecole delle librerie sono in grado di legarsi al vetrino e

l'amplificazione avviene attraverso una PCR detta *Bridge-PCR* (Fig.34). Dapprima, gli oligonucleotidi ancorati funzionano da *primers* e le molecole di libreria a loro ibridate da stampo. In seguito ad un primo ciclo di amplificazione, viene prodotta la prima generazione di copie dei filamenti di DNA, che restano ancorati alla superficie. A seguito della denaturazione, i filamenti stampo tornano di nuovo disponibili per ulteriori processi di ibridazione e allungamento dei *primers* ancorati. Le stesse copie così generate dei filamenti della libreria, funzionano ora da stampo per un'ulteriore amplificazione, mediante ibridazione "a ponte" con gli oligonucleotidi vicini a generare copie aggiuntive in prossimità di quelle iniziali (formazione di un *cluster*). Pertanto, da una singola molecola di DNA di partenza di una libreria, a seguito di cicli di amplificazione successivi vengono generati frammenti clonali di sequenza identica alla prima molecola (possono essere generate più di 1000 copie del frammento iniziale) raggruppati in *cluster* (figura 34).

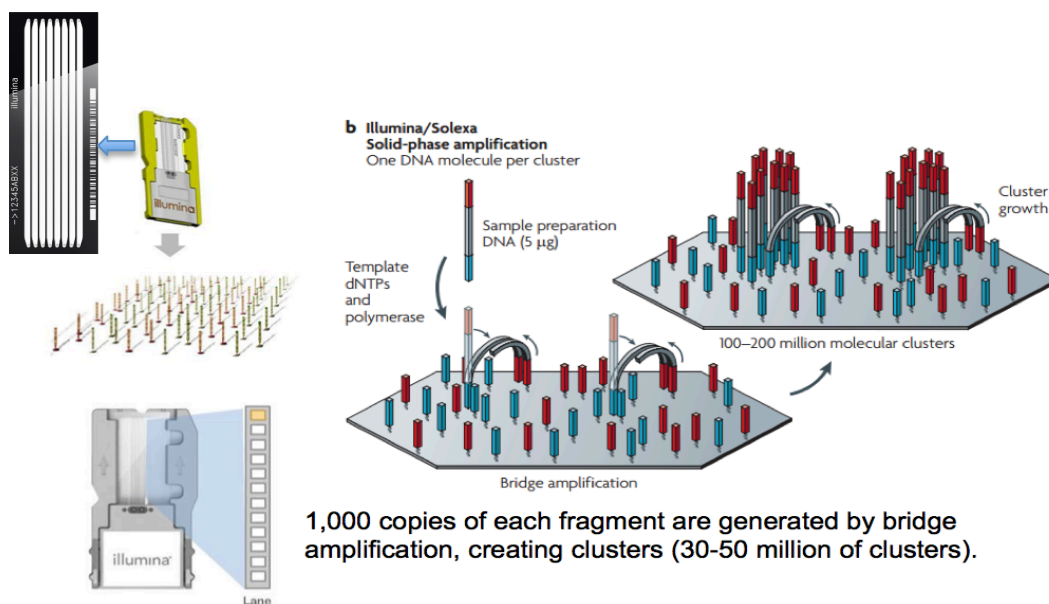


Figura 34: Amplificazione delle librerie Illumina. Sono rappresentate la *flow cell*, il legame dei frammenti alla superficie del *flow cell*; *Bridge-PCR* e i *clusters* degli ampliconi.

La *flow cell* rappresenta anche il supporto della successiva fase di sequenziamento.

3. **Sequenziamento.** Le piattaforme Illumina utilizzano un sistema di sequenziamento mediante sintesi (SBP, *Sequencing by Synthesis*). I *clusters* vengono denaturati e ibridati all'estremità 3' con il *sequencing primer* complementare all'adattatore; ai canali della *flow cell* sono

contemporaneamente aggiunti la DNA polimerasi e tutti e quattro i desossinucleotidi. Questi ultimi sono indicati come nucleotidi *CTR* (*Cyclic Reversible Terminators*) e presentano due caratteristiche fondamentali: ognuno di essi porta legato al 3'-OH un terminatore reversibile della reazione di sintesi, cioè un gruppo bloccante chimicamente eliminabile, così che possa essere incorporato un solo nucleotide per volta; inoltre ognuno di essi è marcato a livello della base azotata con un fluoroforo scindibile diverso per ciascuna base (figura 35). Per ogni ciclo di sequenziamento i 4 dNTP *CTR* competono per essere aggiunti dalla polimerasi alla catena nascente, soltanto quello complementare alla base sul filamento stampo sarà incorporato (figura 35). A seguito dell'incorporazione, il laser eccita il fluoroforo e l'emissione di fluorescenza viene captata e registrata da una camera *CCD*, mentre un computer monitora ogni *cluster* a cui viene assegnata una sequenza (figura 35). Il terminatore reversibile viene quindi staccato dal nucleotide incorporato insieme al fluoroforo. La rimozione del blocco al 3' OH e del fluoroforo prepara ogni filamento all'estensione di un altro nucleotide per rilevare la base seguente [Metzker Michael L., 2010]. I cicli di incorporazione ed emissione di fluorescenza sono impostati dall'operatore in funzione della lunghezza dei frammenti di DNA. Con la piattaforma *MiSeq* possono essere prodotte 25 milioni di *reads* (sequenze) di lunghezza ciascuna 2x300 bp (da entrambe le estremità) e fino a 15 Gb di dati *run*. Si possono ottenere, invece, fino a 1000 Gigabasi per corsa, di lunghezza 2x150 bp con la piattaforma *HiSeq 2500* (figura 36). Ad oggi per l'analisi del DNA antico queste sono le piattaforme maggiormente utilizzate, in quanto consentono il recupero di dati di sequenza relativi a frammenti anche molto corti generando milioni di letture ad un costo contenuto. Un altro vantaggio è la capacità di *multiplexing*, cioè di sequenziare librerie derivanti da campioni diversi nella stessa corsa, grazie alle partizioni fisiche della *flow cell* ed alle librerie dotate di indici campione specifici. In questo modo le librerie di campioni diversi riunite in unico *pool* (insieme di molti campioni) vengono sequenziate insieme (in *multiplex*).

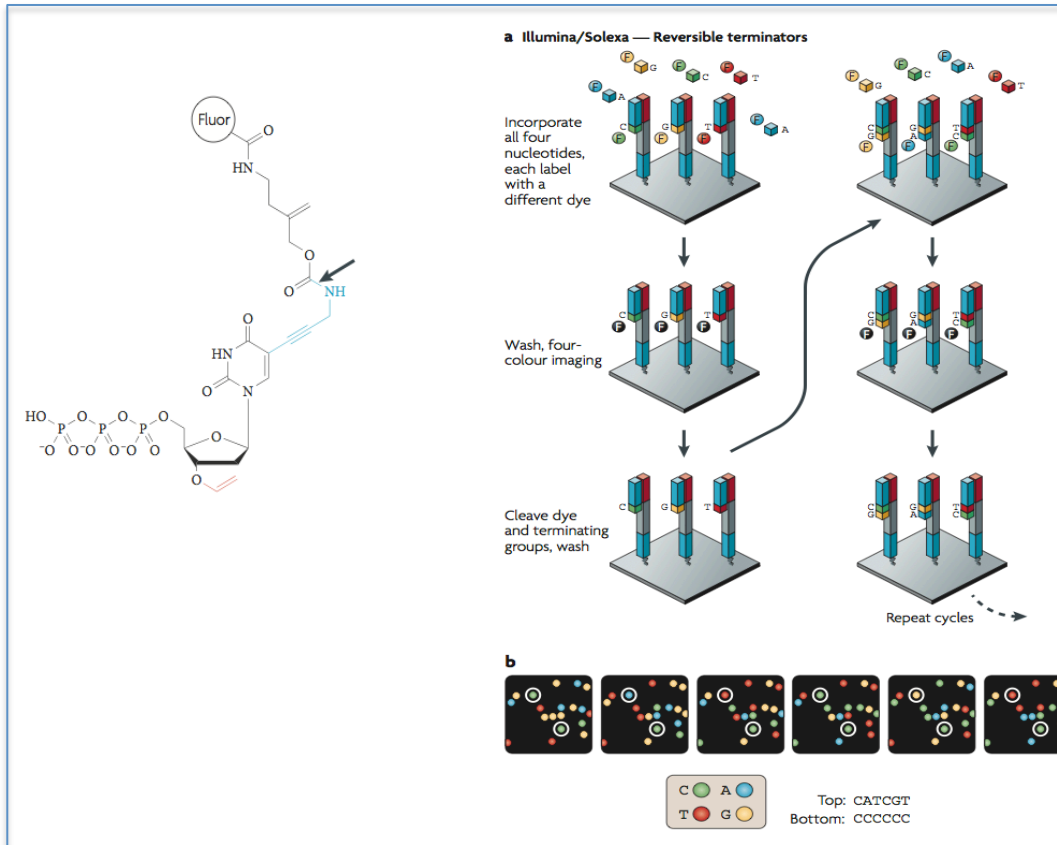


Figura 35: Sequenziamento Illumina. A sinistra un esempio di dNTP *CTR*; a destra **a.** il sistema di *Sequencing by Synthesis*. **b** conversione della luce di emissione in dati di sequenza per due diversi *clusters* [Metzker Michael L., 2010].

						
	MiSeq	NextSeq 500		HiSeq 2500	HiSeq X*	
Key applications	Small genome, amplicon, and targeted gene panel sequencing.	Everyday genome, exome, transcriptome sequencing, and more.		Production-scale genome, exome, transcriptome sequencing, and more.	Population-scale human whole-genome sequencing.	
Run mode	N/A	Mid-Output	High-Output	Rapid Run	High-Output	N/A
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1 or 2
Output range	0.3-15 Gb	20-39 Gb	30-120 Gb	10-180 Gb	50-1000 Gb	1.6-1.8 Tb
Run time	5-65 hours	15-26 hours	12-30 hours	7-40 hours	< 1 day - 6 days	< 3 days
Reads per flow cell†	25 Million‡	130 Million	400 Million	300 Million	2 Billion	3 Billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 125 bp	2 × 150 bp

Figura 36: Caratteristiche delle piattaforme Illumina. Nel presente lavoro di Dottorato sono state utilizzate *MiSeq* ed *HiSeq 2500 High-Output*.

3.9. Sequenziamento mediante piattaforma *MiSeq* dell' mtDNA catturato

I catturati riuniti in un *pool* equimolare pari a 2 nM sono stati sottoposti a sequenziamento ultramassivo. Il sequenziamento delle librerie catturate è stato realizzato presso l'istituto di Tecnologie Biomediche del Consiglio Nazionale delle Ricerche di Milano, su piattaforma *Illumina MiSeq* utilizzando il kit v3 ed impostando 2x75 cicli di sequenziamento (*paired-end*).

3.10. Analisi dei dati di sequenza

Il risultato del sequenziamento è rappresentato da una serie di *files* in formato FASTQ contenenti tutte le sequenze (dette *reads*) prodotte. In particolare, per ciascun campione sequenziato su *MiSeq*, vengono generati due *files*: uno contenente le *reads forward* (R1) ed uno contenente le *reads reverse* (R2) appartenenti alla medesima libreria. Tali sequenze sono state innanzitutto sottoposte a *de-multiplexing*, processo mediante il quale le sequenze di ciascun campione sono identificate e separate dalle altre sulla base della coppia di indici assegnata.

Il FASTQ format è un formato di testo che include sia la sequenza nucleotidica prodotta che la qualità di ogni base chiamata (*score*) ed usa quattro righe per sequenza come di seguito mostrato:

```
@EAS139:136:FC706VJ:2:2104:15343:197393 1:Y:18:ATCACG
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((((***+))%%&&++) (%%&&&).1***-+*''))**55CCF>>>>>>>CCCCCCC65
```

1. La prima inizia con il simbolo @ seguito dal codice identificativo della sequenza;
2. La seconda riporta la sequenza nucleotidica;
3. La terza presenta il simbolo + che funge da separatore, segnalando il termine della seconda riga e quindi della sequenza e l'inizio della quarta riga [Cock et al.2010].
4. La quarta, porta il codice di caratteri ASCII relativo alla qualità della sequenza nella riga 2. Ciascun carattere ASCII corrisponde ad un valore numerico dato dalla somma del valore di qualità (Q) della chiamata di una base più 33 (tabella 3) [Illumina 2011].

Es. il carattere ? corrisponde a 63 dato dalla somma $(Q+33) = (30+33) = 63$

ASCII_BASE=33 Illumina, Ion Torrent, PacBio and Sanger											
Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII
0	1.00000	33 !	11	0.07943	44 ,	22	0.00631	55 7	33	0.00050	66 B
1	0.79433	34 "	12	0.06310	45 -	23	0.00501	56 8	34	0.00040	67 C
2	0.63096	35 #	13	0.05012	46 .	24	0.00398	57 9	35	0.00032	68 D
3	0.50119	36 \$	14	0.03981	47 /	25	0.00316	58 :	36	0.00025	69 E
4	0.39811	37 %	15	0.03162	48 0	26	0.00251	59 ;	37	0.00020	70 F
5	0.31623	38 &	16	0.02512	49 1	27	0.00200	60 <	38	0.00016	71 G
6	0.25119	39 '	17	0.01995	50 2	28	0.00158	61 =	39	0.00013	72 H
7	0.19953	40 (18	0.01585	51 3	29	0.00126	62 >	40	0.00010	73 I
8	0.15849	41)	19	0.01259	52 4	30	0.00100	63 ?	41	0.00008	74 J
9	0.12589	42 *	20	0.01000	53 5	31	0.00079	64 @	42	0.00006	75 K
10	0.10000	43 +	21	0.00794	54 6	32	0.00063	65 A			

Tabella 3: Codice ASCII: caratteri ASCII e corrispondenti valori di qualità (Q) e probabilità di errore (P) [Illumina 2011].

La qualità Q, relativa alla chiamata di ciascuna base nella sequenza è espressa in *PHRED score* legato alla probabilità, P_e che la chiamata della base corrispondente sia errata secondo l'equazione:

$$Q_{PHRED} = -10 \log_{10} (P_e) \text{ e quindi } P_e = 10^{-Q/10}$$

Pertanto, se il *PHRED score* assegna ad una base un valore Q pari a 30 la probabilità che questa base sia stata chiamata in maniera errata è di 1 a 1000 con un'accuratezza del 99,9% come riportato in tabella 4.

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Tabella 4: I valori di *PHRED score* sono correlati in maniera logaritmica alla probabilità di errore e di correttezza della chiamata di una base.

Al fine di garantire una maggiore attendibilità ai risultati ottenuti, le sequenze prodotte dal sequenziamento dell'mtDNA catturato dai 13 campioni in esame, sono state analizzate mediante tre diversi programmi bioinformatici preposti all'analisi del DNA antico e descritti di seguito secondo l'ordine di utilizzo.

3.10.1 Pacchetto FastQprocessing/MIA

Le sequenze sono state inizialmente analizzate mediante un pacchetto di *script* denominati *FastqProcessing*, programmati in linguaggio Python [Kircher 2012]. Questi comandi consentono di estrapolare informazioni sulle sequenze, quali il numero totale, la lunghezza, la qualità del sequenziamento, il contenuto in GC ed il numero di basi non identificate ed indicate con la lettera N. Partendo, poi, dall'assunto che il DNA antico è caratterizzato da frammenti relativamente brevi e che le *reads* contengono l'intera o parte della sequenza degli adattatori, i primi *script* di questo pacchetto consentono di effettuare l'***adapter trimming***, ovvero l'eliminazione delle sequenze degli adattatori. Parallelamente, viene effettuato in questa fase anche il ***merging delle reads***, ovvero la sovrapposizione tra le *reads forward e reverse* (R1 e R2), ottenute dal sequenziamento (*paired-end*). Per realizzare il *merging* è necessario impostare un valore di sovrapposizione (*overlap*) che è deciso in base alla lunghezza delle *reads* generate. Nel nostro caso, il sequenziamento genera *reads* lunghe 76 bp ed è stato impostato un valore di sovrapposizione pari a 10 bp al fine di massimizzare il recupero delle sequenze corte più probabilmente endogene, antiche. Infatti, se la percentuale di *reads* che hanno superato il *merging* è elevata significa che il sequenziamento è di buona qualità, non si sono verificati errori ed i frammenti originari erano di brevi dimensioni, al massimo 142 bp $[(76+76) - 10] = 142$.

Le *reads* (R1 e R2) che si sovrappongono per almeno 10bp, quindi, sono state sottoposte al ***Quality Filtering***, un filtro che permette di eliminare le sequenze aventi un elevato contenuto in GC, tipiche del DNA batterico, e quelle aventi più di cinque basi con un valore di *PHRED score* minore di 15.

A questo punto le *reads merged* che hanno superato il *Quality Filtering*, vengono isolate da tutte le altre che non hanno risposto ai criteri impostati e vengono utilizzate per il mappaggio sul genoma mitocondriale. A questo scopo è stato utilizzato il ***Mapping Iterative Assembler (MIA)***, composto da una serie di *script* in grado di allineare tutte le *reads* appartenenti ad una determinata libreria con una sequenza di riferimento, creando una sequenza consenso [Green et al., 2008]. Questo programma è stato appositamente realizzato per l'analisi del DNA antico; infatti per l'allineamento e la creazione della sequenza consenso, utilizza una matrice di sostituzione posizione-specifica che tiene conto dei possibili danni presenti sul DNA antico. Inoltre, poiché le fasi sperimentali dell'analisi

dell'aDNA, come descritto, prevedono diversi passaggi di amplificazione per aumentarne la quantità, per ogni libreria di ciascun campione sono state generate molteplici copie ed il primo *script* MIA è in grado di collassare tutte le copie di una stessa sequenza in un'unica *read* consenso, richiamando in caso di *mismatch* la base con il più alto valore di qualità. Questo è importante perché un alto numero di *reads* identiche può portare a stime errate della frequenza di polimorfismi. L'*output* dello script è rappresentato da due o più *file* di mappaggio *unique* (con estensione *.maln*), contenenti le *reads* copie collassate della stessa sequenza, e da due o più *file not-unique* (con estensione *.maln*), nei quali le *reads* clonali sono state mappate mantenendole separate. I *file unique* e *not-unique* vengono poi utilizzati come *input* per i successivi *script*, al fine di generare la sequenza consenso mappandoli sulla sequenza di riferimento *rCRS* (*Revised Cambridge Reference Sequence*) depositata in *GenBank* (accession number: NC_012920) [Andrews et al. 1999]. Il rapporto fra *reads unique* e *not-unique* permette di definire il *cluster factor*, un valore che fornisce una stima della profondità del sequenziamento di ciascun campione. Inoltre, lo *script MapAssembler* processa le *reads unique* restituendo molteplici informazioni sul mappaggio effettuato:

- *average coverage*, la profondità del sequenziamento per ciascuna base, cioè quante volte ciascuna base è stata chiamata in quella posizione nelle *reads*;
- *average length*, la lunghezza media dei frammenti;
- *fold coverage*, la copertura di ogni base, ovvero sul totale delle 16.569 basi del mtDNA umano, quante basi sono coperte almeno 1, 2, 3, 4, e 5 volte. Dato questo che ci permette di conoscere la percentuale di copertura dell'mtDNA e se vi sono posizioni non coperte, non sequenziate.

3.10.2 SeqPrep

Si tratta di un insieme di *script* eseguiti manualmente dall'operatore, indicati nel complesso pacchetto SeqPrep appositamente sviluppati per l'analisi del DNA antico. Come con il pacchetto precedente anche con SeqPrep le prime fasi dell'analisi delle sequenze sono rappresentate dall'eliminazione delle sequenze degli adattatori e dalla sovrapposizione tra le *reads forward* e *reverse* (R1 e R2) ottenute dal sequenziamento (*paired-end*). Per effettuare il *merging* delle *reads* è impostato il valore di sovrapposizione pari a 10 bp così da scartare tutte le *reads* di lunghezza superiore a 140 bp. Inoltre, il programma elimina anche le sequenze troppo brevi, di lunghezza inferiore a 30 bp in quanto potrebbero mappare in maniera aspecifica. Le *reads merged* vengono, pertanto, mappate sulla sequenza di riferimento per il mtDNA umano, la *rCRS*, mediante il *software Burrows-*

Wheeler Aligner, BWA [Li e Durbin 2009; Li e Durbin 2010]. Quest'ultimo mappa le *reads merged* sulla sequenza di riferimento mitocondriale *rCRS* e restituisce come *output* un *file* con estensione *SAM* (*Sequence Alignment/Map*) che può essere inserito nel programma *MapDamage*, il *framework* in grado di determinare il *pattern* di *misincorporazione*. *BWA* è un pacchetto di *software* costituito da tre algoritmi *BWA-backtrack*, *BWA-SW* e *BWA-MEM*. Il primo algoritmo è stato scritto specificamente per *reads*, generate dalla piattaforma Illumina lunghe fino a 100 bp; gli altri due invece per sequenze di lunghezza fra 70bp e 1Mbp. *BWA-MEM* è l'algoritmo disegnato più di recente e data la sua velocità e accuratezza è il più consigliato per analisi di alta qualità, ha una resa migliore rispetto a *BWA-backtrack* per le *reads* di 70- 100bp.

Per migliorare l'efficienza di mappaggio sono stati impostati determinati parametri, quali la tolleranza al massimo di due *gaps*. I *file* di *output* del mappaggio sono poi stati sottoposti alla rimozione dei duplicati prodotti durante le fasi di arricchimento, mediante *SAMtools rmdup* [Li et. al, 2009]. Le sequenze a questo punto sono state analizzate per la selezione della qualità delle basi:

vengono selezionate le basi aventi una buona copertura mentre quelle non rispondenti ai valori di qualità sono indicate con N. Le *reads* con una buona qualità di mappaggio sono utilizzate per generare la sequenza consenso mediante *VCFgenome software* [P. Danecek et al., 2011]

I *files* con estensione *SAM* possono ora essere letti da *MapDamage*.

3.10.3 Software Eager

Sviluppato nel 2016, questo programma consente di effettuare in maniera più semplice e rapida l'analisi di sequenze derivanti da varie piattaforme di sequenziamento. Non richiede, infatti, particolari conoscenze di linguaggio bioinformatico e svolge automaticamente ed in contemporanea le varie fasi dell'analisi, che con gli altri *software* vengono eseguite manualmente dall'operatore in successione e fornisce come *output* un *file* riassuntivo di tutti i risultati ottenuti. E' in grado di fare questo grazie all'interfaccia grafica di cui è dotato: *Graphical User Interface*, (*GUI*) in cui l'operatore seleziona i *file* di *input* in formato *FASTQ* relativi alle *reads forward* e *reverse* di ciascun campione, la sequenza di riferimento *rCRS* ed imposta i parametri di mappaggio e di calcolo del *pattern* di *misincorporazione*. Da questa interfaccia grafica è possibile accedere anche ai *software BWA* e *DeDup* utilizzati nell'analisi. La prima fase del *workflow Eager* consiste nel *Preprocessing*: analisi dei *files* in formato *FASTQ* e valutazione della qualità dei dati ottenuti dal sequenziamento. Per fare ciò utilizza il metodo *Clip and Merge* con il quale rimuove le sequenze degli adattatori alle estremità dei frammenti ed esegue il *merging* delle *reads*. Per la fase di

mappaggio, poi, allinea le *reads* ottenute nella fase precedente, alla sequenza consenso di riferimento utilizzando *BWA* [Li e Durbin 2009; Li e Durbin 2010]. I *files* di *output* in formato SAM vengono sottoposti alla rimozione dei duplicati generati nelle fasi sperimentali di amplificazione delle librerie, mediante *DeDup*. Infine il risultato del mappaggio viene sottoposto a *QualiMap* [Okonechnikov et al., 2016].

L'ultima fase dell'analisi delle sequenze ottenute è stata loro **autenticazione**. Questa è stata realizzata mediante l'analisi del pattern di *misincorporazione* alle estremità dei frammenti con *software MapDamage* e l'analisi del pattern di *contaminazione* utilizzando il *software ContamMix* di seguito descritti.

3.10.4 Analisi del *pattern* di *Misincorporazione*

Il *pattern* di *misincorporazione* è stato valutato attraverso lo strumento bioinformatico *MapDamage 2.0*. Si tratta di un *framework* scritto in Python ed R in grado di misurare lo stato di degradazione del DNA, in particolare rileva e quantifica i *pattern* di danneggiamento all'interno dei dati di sequenza di aDNA generati dalle piattaforme *NGS* [Ginolhac et al. 2011; Jöhnsson et al. 2013]. Come descritto nel capitolo introduttivo del presente lavoro, a causa dei fenomeni di deaminazione post-mortem della citosina, nelle molecole di DNA antiche si registra un aumento delle transizioni C → T al 5' della *read* ed un corrispondente aumento della sostituzione G → A al 3' [Briggs et al., 2007]. *MapDamage 2.0* si fonda sull'assunto che gli eventi di danneggiamento dipendano solo dalla posizione sulla sequenza e dalla deaminazione post-mortem, rifacendosi al modello statistico descritto da Briggs che usa come fattori chiave del danno le rotture del filamento, le estremità sporgenti, la deaminazione della citosina [Briggs et al., 2007]. *MapDamage*, realizza una stima statistica di tipo bayesiano che considera quattro parametri fondamentali: la probabilità di terminazione di una sporgenza (λ), la probabilità di deaminazione della citosina all'interno del doppio filamento (δ_D), la probabilità di deaminazione della citosina all'interno del singolo filamento (δ_S); il tasso di differenza media non causato da danni al DNA tra la sequenza di riferimento e quella del campione sequenziato (θ) [Jöhnsson et al. 2013].

3.10.5 Analisi del *pattern* di *Contaminazione*

L'analisi delle contaminazioni è stata effettuata utilizzando il *software ContamMix*, il quale mediante un approccio statistico di tipo bayesiano calcola la

probabilità che il profilo di DNA ottenuto da ciascun campione derivi da un unico individuo [Fu et al., 2013]. Questo strumento bioinformatico utilizza i *file* di *output* generati con *BWA* calcolando la probabilità che il profilo di DNA ottenuto provenga da un solo individuo in base a due parametri:

- Il mappaggio delle *reads merged* del campione sulla sequenza consenso generata;
- L'allineamento della sequenza consenso del campione alle sequenze di DNA mitocondriale di 311 individui moderni che rappresentano potenziali contaminanti.

In questo modo stima con un intervallo di confidenza del 95% (C.I. 95) la frazione di *reads* che mappano con un punteggio più elevato sulla sequenza consenso ricostruita, piuttosto che su quelle dei potenziali contaminanti. In particolare, *ContamMix* prende in considerazione le *n reads* del campione, la sequenza consenso generata per il campione, *m* (311) genomi mitocondriali moderni, considerando che le *reads* siano state ottenute da *m+1* genomi. Per ogni *read* e per ogni genoma sono definiti tre parametri: numero di basi della *read* presenti nel genoma; numero di basi della *read* non presenti nel genoma; inserzioni e delezioni della *read* presenti (1) o non presenti (0) nel genoma. Il *software* poi, scarta le *reads* contenenti inserzioni e delezioni che non sono presenti in alcun genoma, assumendo che siano state generate da errori durante il processamento delle sequenze. Definito, poi, il tasso di errore come la probabilità che per una singola base avvenga un errore di sequenziamento, viene usata una classe di algoritmi MCMC (*Monte Carlo Markov Chain*) al fine di stimare la probabilità che la contaminazione possa derivare da un numero arbitrario di individui ($m > 1$).

3.10.6 Identificazione dell'Aplogruppo mitocondriale

Come descritto in precedenza, gli alleli del genoma mitocondriale (mtDNA) sono sempre associati a formare aplotipi, ove per aplotipo s'intende la combinazione degli alleli in loci associati, tra i quali cioè, non avvengono fenomeni di ricombinazione.

Aplotipi differenti sono generati a partire da un aplotipo ancestrale per effetto di variazioni nucleotidiche ai singoli loci. Per aplogruppo s'intende, invece, un insieme di aplotipi accomunati da un *set* di loci polimorfici recanti le medesime varianti alleliche ed ereditati dal un progenitore comune. Per l'identificazione degli aplogruppi è stato utilizzato il *software online HaploFind*: una piattaforma bioinformatica che sfrutta un algoritmo in grado di classificare nuove sequenze di DNA mitocondriale in base alla nomenclatura *PhyloTree* [Vianello et al., 2013]. Il programma confronta la sequenza di riferimento *rCRS* del genoma mitocondriale, con la sequenza consenso generata per il campione in esame e, a seguito del

confronto, ne individua i polimorfismi assegnando l'aplogruppo di appartenenza. *HaploFind* attribuisce anche un punteggio (*score*) che esprime l'affidabilità dell'aplogruppo assegnato: quanto è probabile che l'aplogruppo assegnato ad un campione sia effettivamente quello di appartenenza.

3.11 Fasi sperimentali realizzate al Dipartimento di Genetica di Harvard

Le librerie *half-UDG* dei 13 campioni in esame, sono state inviate al laboratorio del Prof. David Reich del dipartimento di Genetica dell'Università di Harvard, in modo da consentire la replicazione dell'analisi del DNA mitocondriale. Date le notevoli criticità presentate da questi campioni, in condizioni precarie di conservazione per il DNA endogeno, la replicazione dell'analisi del mtDNA si è resa auspicabile al fine di dare maggior robustezza ai risultati ottenuti.

Presso il medesimo dipartimento sono state effettuate le **analisi sul genoma nucleare** di tali campioni. In particolare è stato realizzato un sequenziamento **shotgun a bassa copertura** per la determinazione di genere. Per il protocollo utilizzato vedasi le seguenti note bibliografiche [Skoglund et al. 2013; Fu Q. et al. 2016].

Inoltre, dato che il gruppo del Prof. Reich ha ideato e messo a punto la metodica per la cattura di polimorfismi a singolo nucleotide in genomi nucleari antichi mediante la costruzione di sonde capaci di ibridare a livello di determinati *SNPs* in molecole di DNA con i caratteristici danni C->T al 5', è stata realizzata anche la **cattura ed il sequenziamento su piattaforma HiSeq2500 di 1.240.000 SNPs dei genomi nucleari dei campioni**. Per ogni *SNP* sono state disegnate quattro sonde di 52 nucleotidi ciascuna, di cui due capaci di ibridare le regioni immediatamente fiancheggianti lo *SNP* da entrambe le estremità e due centrate sullo *SNP* e contenenti l'uno oppure l'altro allele alternativo, rispettivamente [Qiaomei Fu et al. 2015; Qiaomei Fu, Meyer M., et al. 2013]. In allegato al presente lavoro, l'elenco degli *SNPs* considerati. Il protocollo della cattura utilizzato è indicato come *Solution-phase Hybrid Selection, SHS*, il cui schema è riportato in figura 37. Sonde di DNA biotinilate sono poste in soluzione insieme alle librerie genomiche indicizzate *half-UDG* precedentemente preparate. Le sonde biotinilate libere in soluzione ibridano selettivamente le porzioni complementari delle librerie e l'*heteroduplex* sonda-frammento viene catturato da biglie magnetiche ricoperte di streptavidina aggiunte alla miscela [Gasc C., Eric Peyretailade et al. 2016]. I frammenti della libreria non ibridati alle sonde sono eliminati dalla soluzione in seguito a lavaggi successivi, mentre i frammenti

ibridati sono eluiti dalle biglie e amplificati per il sequenziamento [Gnirke A. et al., 2009]. Per il dettaglio del protocollo sperimentale eseguito si rimanda a note bibliografiche [Fu Q et al 2015; *supplementary notes* da Gasc C. 2016].

Le analisi dei catturati nucleari sono tuttora in corso, ma in questo lavoro presento i primi risultati alla sezione successiva.

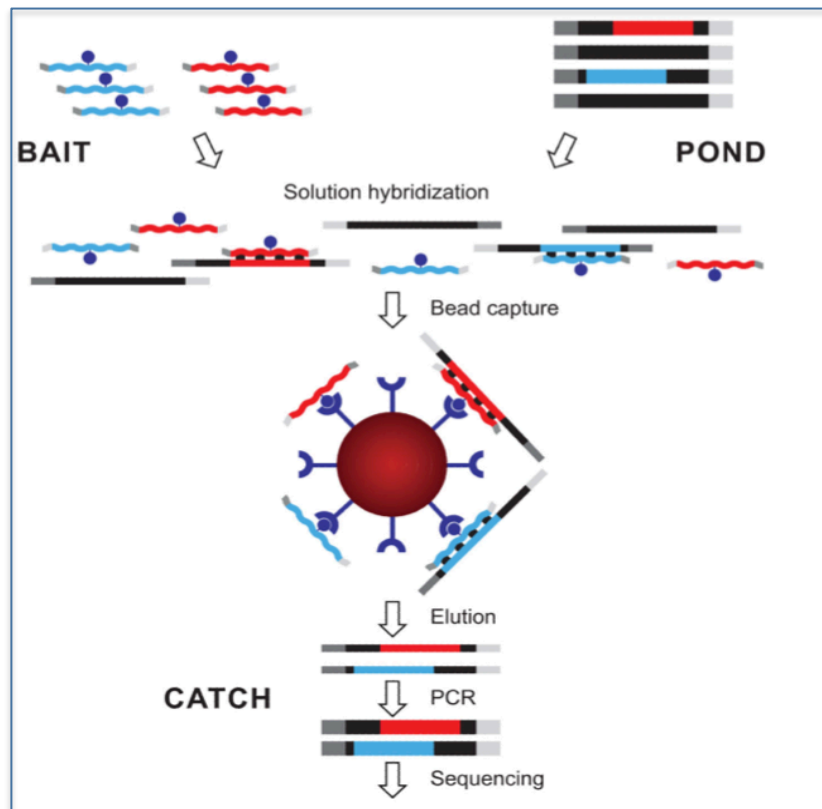


Figura 37: Schema della cattura degli *SNPs* mediante Selezione dell'ibrido in Soluzione (*SHS*). Spiegazione nel testo. Immagine modificata da [Gnirke A. et al., 2009].

4. Risultati

I diciotto campioni in esame, sedici frammenti ossei e due reperti dentali, recuperati in fase di restauro di alcuni calchi umani nell'area dell'*Insula Occidentalis* a Pompei, sono stati sottoposti a pulizia e produzione della polvere d'osso presso il Laboratorio di Antropologia molecolare dell'Università degli Studi di Firenze. Pertanto, in ciascun campione è stato praticato un foro per ottenere una quantità di polvere sufficiente alla successiva fase di estrazione del DNA. Il protocollo di estrazione utilizzato, messo a punto per ottimizzare il recupero di frammenti corti di DNA, quindi congeniale all'estrazione di molecole antiche altamente degradate, richiede una quantità minima di 50 mg di polvere d'osso al fine di un'estrazione efficiente [Dabney et al., 2013]. A causa dell'elevato stato di degradazione dei frammenti ossei, da cinque campioni non è stato possibile ricavare la quantità minima di polvere. In tabella 5 sono riportati tutti i campioni con indicazione di quelli che non hanno proseguito nella fase di estrazione.

Individuo	Sito di recupero	Campione
Adulto	Casa del bracciale d'oro	C-21
Adulto	Casa del bracciale d'oro	C-34
Infante	Casa del bracciale d'oro	C-35
Infante	Casa del bracciale d'oro	C-23
Adulto	Casa del bracciale d'oro	C-32
Adulto	Casa del bracciale d'oro	C-33 (dente)
Adulto	Casa del bracciale d'oro	C-8_NO
Adulto	Casa del bracciale d'oro	C-6
Adulto	Casa del Bracciale d'oro	C-7_NO
Adulto	Santuario	C-86
Adulto	Granai del foro	C-27_NO
Adulto	Casa del criptoportico	C-11
Adulto	Terme <u>stabiane</u>	C-77
Adulto	Terme del <u>sarno</u>	C-66_NO
Adulto	Terme del <u>sarno</u>	C-67
Adulto	<u>Macellum</u>	C-30
Adulto	<u>Macellum</u>	C-31 NO
Adulto	Villa dei Misteri	C-5 (dente)

Tabella 5: Campioni in esame con indicazione di quelli che non hanno proseguito nella fase di estrazione.

Il DNA estratto dai restanti tredici campioni è stato quindi utilizzato per un volume pari a 20µl per preparare le librerie genomiche *half-UDG* come descritto [Meyer and Kircher, 2010] [N. Rohland, E. Harney 2014]. Per ciascun campione e controllo negativo (k- estrazione e k- libreria) è stata scelta la combinazione di una coppia di indici diversa al fine di identificare in maniera univoca le sequenze di ognuno di essi a seguito del sequenziamento. Le librerie così prodotte una volta arricchite per amplificazione con il *kit Herculanase* sono state quantificate mediante *Agilent BioAnalyzer 2100*. In figura 38 sono riportati a titolo di esempio i profili ottenuti per alcuni dei campioni in esame. L'*Agilent BioAnalyzer2100* consente di effettuare un'elettroforesi quantitativa delle molecole in esame. Pertanto, restituisce per ciascuna libreria un grafico in cui è posta la concentrazione del DNA espressa in unità di fluorescenza (FU) in funzione della lunghezza dei

frammenti espressa in coppie di basi ($Y=f(X)$). La corsa elettroforetica è realizzata accanto ad un marcatore a concentrazione nota e di lunghezza in bp nota e sul grafico sono sempre evidenti i suoi picchi minore e maggiore, rispettivamente di 15bp e 1500bp. L'integrazione dell'area sotto la curva quantifica la concentrazione del DNA.

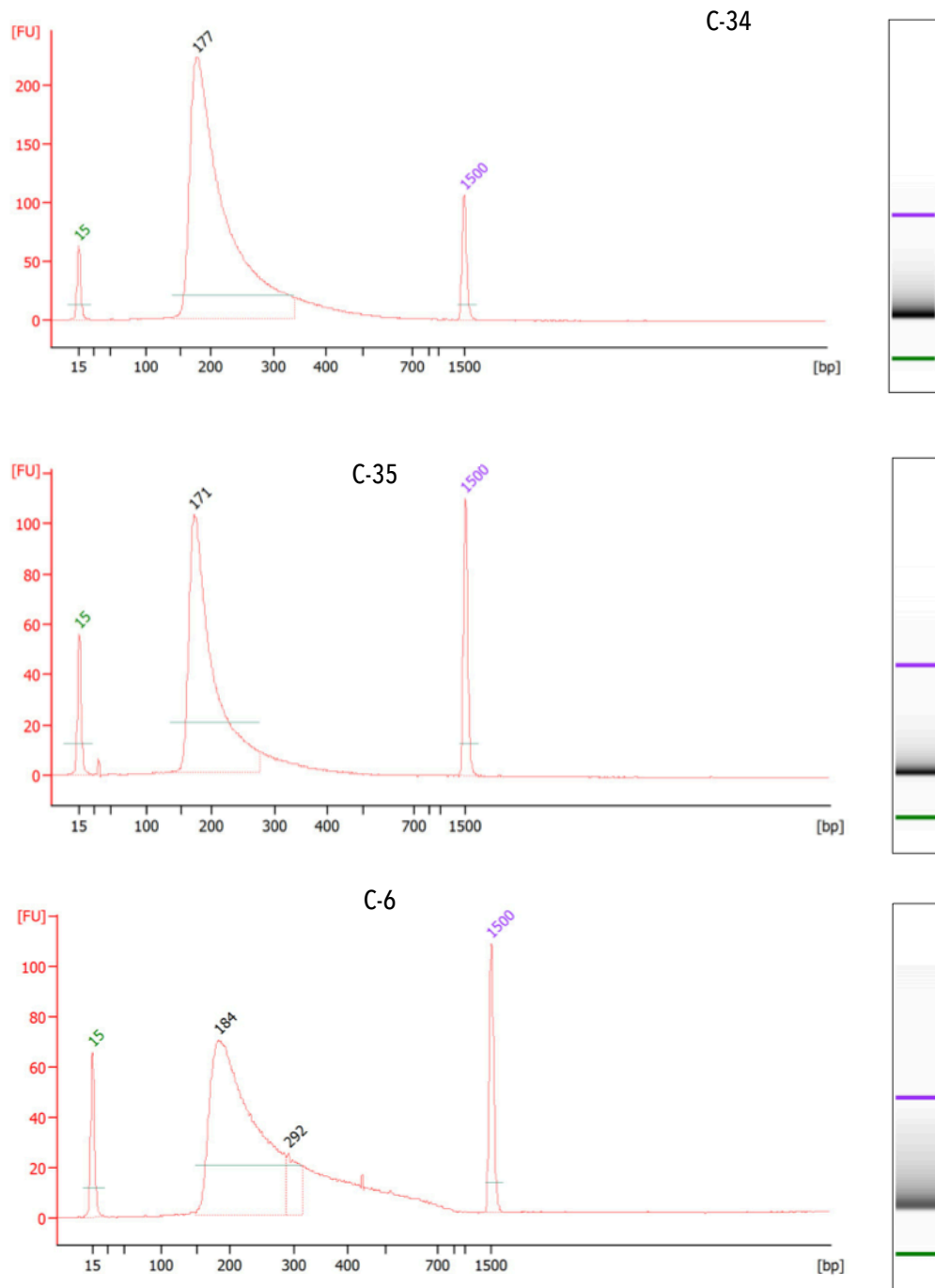


Figura 38: Elettroforesi quantitativa realizzata mediante *Agilent Bioanalyzer2100* delle librerie di tre campioni.

Già con la visualizzazione del profilo delle librerie ottenute all'*Agilent*, è possibile ricavare le prime informazioni circa l'autenticità delle molecole di DNA endogeno in analisi. Infatti, le librerie prodotte da molecole antiche altamente degradate, mostrano all'*Agilent* una lunghezza compresa tra le 170 e le 350bp, con una media vicina alle 200bp compatibile con frammenti di 140bp circa, legati agli adattatori e agli indici ad entrambe le estremità. I profili delle librerie ottenute dai campioni in analisi, mostrano un picco principale tra 171 e 184 bp, ed alcuni un ulteriore picco più basso a maggiore lunghezza intorno alle 300bp che potrebbe indicare la presenza di DNA contaminante moderno o la formazione di *heteroduplex* generati dall'appaiamento di basi complementari di frammenti diversi.

Le librerie a questo punto sono state utilizzate per le analisi successive sul DNA mitocondriale e nucleare.

4.1 Risultati dell'analisi del DNA mitocondriale

Le librerie genomiche *half-UDG* prodotte e quantificate presso il laboratorio dell'Università di Firenze, sono state qui utilizzate per la cattura del DNA mitocondriale seguendo il protocollo pubblicato da Maricic e colleghi nel 2010 [Maricic et al 2010]. Le medesime librerie sono state inviate al laboratorio del Prof. Reich, in Boston, che ha replicato le catture del mtDNA, il sequenziamento e la relativa analisi dei dati di sequenza. Di seguito riporto i risultati ottenuti dall'analisi del genoma mitocondriale realizzata presso il laboratorio di Firenze e che sono del tutto sovrapponibili a quelli ottenuti dai colleghi Americani.

Per procedere alla cattura del DNA mitocondriale delle librerie, abbiamo stimato la concentrazione delle librerie all'*Agilent* nell'intervallo di lunghezze comprese tra 150-400 bp. Passaggio, questo, fondamentale per un'efficiente cattura: i campioni inseriti nella stessa reazione devono essere fra loro equimolari. La concentrazione delle due sonde, invece, è stata determinata al *NanoDrop* ed è stata stimata per la sonda Mt1= 128,80 ng/μl e per la sonda Mt2= 112,5ng/ μl. Tali concentrazioni sono state utilizzate per stabilire il volume di sonde necessario ad ottenere una concentrazione (per sonda) pari a 250 ng (500 ng totali) secondo il protocollo di Maricic. Della libreria di ciascun campione posto nella miscela di reazione, a seguito della cattura, soltanto il DNA mitocondriale umano ibridato alle sonde viene eluito e quindi recuperato per poter essere sequenziato. L'eluito è stato purificato ed amplificato mediante il *kit Herculanase* al fine di raggiungere la concentrazione ottimale per il sequenziamento, pari a 1×10^{12} copie/μl.

I catturati riuniti in una miscela equimolare pari a 2 nM, sono stati sottoposti a sequenziamento ultramassivo su piattaforma *Mi-Seq (Illumina)* presso l'istituto di tecnologie biomediche del CNR di Milano.

Il sequenziamento ha fornito come *output* per ciascuno dei tredici campioni due *files* uno contenente le *reads forward* (R1) e l'altro le sequenze *reverse* (R2) le quali sono state individuate e separate dalle altre, sulla base della coppia di indici assegnata. Le sequenze sono state quindi analizzate mediante i programmi bioinformatici FastQprocessing/MIA, SeqPrep ed Eager ottenendo risultati analoghi; per praticità mi riferirò al primo *software*. Il pacchetto *FastQprocessing* effettua contemporaneamente la rimozione delle sequenze degli adattatori e la sovrapposizione delle *reads* R1 ed R2 per una regione di 10 bp. Da questa prima analisi si ottengono informazioni riguardanti il numero originario di *reads* R1 ed R2, il numero totale di *reads* generate per entrambi i filamenti, le *reads* che non hanno superato il *merging* (*total#separated reads*) perché hanno disatteso i parametri di sovrapposizione impostati, e la percentuale di *merging* definita dal rapporto tra le *merged reads* e le *reads* totali. Relativamente alla percentuale di *merging*, quasi tutti i campioni ad eccezione del C6, C30, C11 e C67 hanno mostrato valori oscillanti tra il 95 e il 96%, rientrando, quindi, nell'intervallo ottimale, 61,85% - 95,4% stabilito per i frammenti corti [Kircher M. 2012]. Quindi le *reads merged* sono state sottoposte a *quality filtering* per scartare tutte quelle sequenze che non rispettano i criteri di qualità impostati. Questo “filtro di qualità” restituisce come *output* una percentuale di *merged quality filtered reads*. Quasi tutti i campioni, ad eccezione di C6, C30, C11 e C67, mostrano una percentuale di *reads merged quality filtered* superiore al 95%, ad indicare che il sequenziamento è andato a buon fine: la maggior parte delle sequenze che hanno superato il *merging* hanno superato anche il *quality filtering*.

A questo punto, le sequenze *merged* filtrate sono state mappate sulla sequenza mitocondriale di riferimento, la *rCRS*, (Andrews et al., 1999) utilizzando il *software Mapping Iterative Assembler* (MIA), un pacchetto di *script* specifico per il DNA degradato che tiene conto dei possibili danni che il DNA antico può presentare. Il primo *script* ha collassato tutte le *reads* allineate con la medesima direzione e con le stesse coordinate di inizio e fine in un'unica sequenza, dato che esse sono i cloni originati dalla piattaforma *Illumina* durante la fase di amplificazione per formare i *clusters* dei vari frammenti. In questo modo ha generato i frammenti unici, indicati come *fragments unique*; inoltre, ha allineato tutti i frammenti separati e quindi non unici, *fragments not unique*. Il rapporto fra sequenze *not unique* ed *unique* genera il *cluster factor*: un valore numerico che indica quante copie di ogni frammento sono state in media sequenziate e fornisce una stima del grado di profondità del sequenziamento. I valori compresi tra 2 e 10 sono considerati ottimali, valori inferiori indicano un sequenziamento poco profondo, cioè i frammenti catturati non sono stati sequenziati in maniera

ottimale, mentre valori superiori indicano un sequenziamento troppo profondo con molte informazioni ridondanti. I campioni in esame hanno mostrato per la maggior parte un *cluster factor* che rientra nell'intervallo ottimale, ad eccezione dei campioni C6, C30, C11 e C67 che hanno riportato valori superiori. Il *software MIA* processa le *reads unique* in maniera tale da stimare anche la copertura media dell'intero mitocondrio, definita *Average Coverage* calcolata in base a quante sequenze uniche coprono ogni singola base del DNA mitocondriale. In altre parole maggiore è il valore di questo parametro, maggiore sarà l'attendibilità della sequenza consenso ricostruita per ogni campione. I valori di copertura mitocondriale ottenuti per i 13 campioni in esame sono riportati in **tabella 6**. La copertura mitocondriale media risulta elevata per i primi sei campioni in tabella, eccetto per il C32.

A questo punto veniamo al nodo centrale dell'analisi dei dati da DNA antico: valutare se le sequenze ottenute hanno le caratteristiche distintive del DNA antico e quindi stimare la loro autenticità. A questo scopo è stato valutato il *pattern* di *misincorporazione* mediante il *software MapDamage 2.0* e la presenza di DNA contaminante mediante il *software ContamMix*.

Come descritto in precedenza, il programma *MapDamage 2.0* [Ginohlac et al., 2011] è stato appositamente sviluppato per stimare il tasso di *misincorporazioni* nelle sequenze generate da piattaforme *NGS*. Ricordiamo che il principale danno *post-mortem* a carico delle basi azotate, che identifica come autentica una molecola di DNA antica, è la deaminazione della Citosina ad Uracile, che porta in fase di sequenziamento alla transizione C→T e quindi sul filamento complementare alla sostituzione G →A [Briggs et al.,2007] [Brotheron et al., 2007]. E' stato sperimentalmente riscontrato che la frequenza di queste sostituzioni C→T e G→A è maggiore alle estremità sporgenti a singolo filamento dei frammenti che si generano in seguito ai fenomeni degradativi, piuttosto che nelle porzioni centrali a doppio filamento. Al contrario, nelle sequenze di DNA moderno questo *pattern* non è visibile oppure è presente in una percentuale inferiore al 5% [Briggs et al. 2007]. Pertanto, il *software MapDamage 2.0* allinea la sequenza consenso ottenuta per ciascun campione alla sequenza di riferimento *rCRS* andando a rilevare i nucleotidi alle estremità 5' e 3' dei frammenti (il primo ntd, l'ultimo ntd e nelle regioni interne delle estremità). In base ai nucleotidi rilevati alle estremità 5' e 3' il programma elabora le frequenze di *misincorporazione*, cioè le frequenze degli appaiamenti errati col genoma di riferimento. E' stato dimostrato che la frequenza delle sostituzioni C→T al 5' è correlata con l'età del reperto: per campioni con più di 500 anni le librerie non trattate con l'enzima UDG hanno un tasso di tali sostituzioni superiore o uguale al 10% [Sawyer et al., 2012]. Nel caso di librerie *half-UDG* da campioni antichi, il tasso di tali sostituzioni è più basso ed è stato stimato maggiore o uguale al 3% [Rohland N., 2015]. Pertanto, con questo tipo di librerie le molecole di DNA

endogene antiche devono presentare un tasso di questo danno $\geq 3\%$. I risultati ottenuti per i 13 campioni in analisi sono riportati nella seconda colonna della tabella 6. Come si osserva, presentano tutti un tasso di danno al 5' superiore alla soglia prevista per le molecole antiche, ad eccezione dei campioni C34 e C67.

Nella quinta colonna della tabella è riportato il risultato di un'ulteriore valutazione dell'autenticità delle molecole antiche in esame. Mediante il *software ContamMix*, sono state allineate le *reads merged* ottenute per ogni campione alla sequenza consenso del campione, al fine di valutare il tasso di appaiamento alla consenso nei casi in cui c'è stata abbastanza copertura mitocondriale media da ottenere una consenso [Green et al., 2008]. Per ciascun campione, il tasso di appaiamento di *reads* endogene alla consenso deve essere compreso tra il 95 ed il 100% [Rohland N., 2015] [Green et al., 2008]. Come si osserva dalla tabella, per i campioni C-67, C-11, C-30 e C-6 non è stato possibile costruire una sequenza consenso a causa della copertura mitocondriale troppo bassa, negli altri casi, gli appaiamenti alla consenso sono compresi nell'intervallo previsto per *reads* endogene, ad eccezione dei campioni C-32, C-77, C-86, C-21.

Il *software ContamMix* mediante un'analisi statistica bayesiana, calcola la probabilità che il profilo di DNA ottenuto per ciascun campione appartenga ad un singolo individuo con un intervallo di confidenza del 95%, in tabella 6 i risultati per ContamMix.95. Questo programma, per ciascun campione prende in considerazione, come descritto, l'allineamento delle *reads merged* alla sequenza consenso ricostruita del campione e l'allineamento della sequenza consenso del campione a 311 sequenze di mtDNA di potenziali contaminanti moderni. In tal modo stima il numero di *reads* che mappano con un punteggio più elevato sulla sequenza consenso piuttosto che su quelle dei potenziali contaminanti con un intervallo di confidenza del 95%. Come si osserva dalla tabella, i campioni C5, C34, C35, C33, C23, mostrano valori compresi tra il 97 ed il 100%, mentre gli altri, mostrano valori nettamente inferiori ad indicare con alta probabilità la presenza di sequenze contaminanti moderne.

Infine, la sequenza consenso ricostruita per ogni campione è stata utilizzata per la determinazione dell'aplogruppo mitocondriale di appartenenza mediante il *software online HaploFind*. Questo programma identifica i polimorfismi presenti nella sequenza consenso del campione, scritta in formato fasta, rispetto alla sequenza di riferimento *rCRS*, associandone l'aplogruppo. Gli aplogruppi identificati per i campioni in esame sono riportati in tabella 6. Il programma, inoltre, associa un punteggio all'aplogruppo assegnato ad indicare la probabilità che questo sia effettivamente l'aplogruppo di appartenenza. Nel caso dei campioni a più alta copertura mitocondriale, il punteggio dell'aplogruppo assegnato risulta elevato, tra 0.90 e 1.00; mentre i campioni aventi una minore copertura

mitocondriale, C32, C77, C86, C21 hanno avuto un punteggio più basso (tra 0.50 e 0.70). Ad ogni modo gli aplogruppi dei nostri campioni sono risultati tutti differenti.

Sample name	Damage in first nucleotide (should be >0.03)	mt.cov	mt.haplogroup	mtDNA match rate to consensus (contamMix)	ContamMix.95 CI	nuclear.capture
C 5	0,10	117	H	1	[0.995,1.000]	Promote.to.1240k
C 32	0,10	5	K1	0,845	[0.797,0.880]	Failed.screening
C 34	0,02	111	U1a	0,996	[0.993,0.998]	Promote.to.1240k
C 35	0,04	14	T2c	0,983	[0.978,0.987]	Promote.to.1240k
C 33	0,09	70	N1b1a	0,991	[0.987,0.995]	Promote.to.1240k
C 23	0,04	170	H	0,998	[0.995,1.000]	Promote.to.1240k
C 77	0,04	8	R	0,524	[0.476,0.563]	Failed.screening
C 67	0,01	1	Failed.screening
C 86	0,08	2	HV	0,406	[0.048,0.719]	Failed.screening
C 11	0,05	1	Failed.screening
C 21	0,06	2	H1h1	0,403	[0.042,0.711]	Failed.screening
C 30	0,09	1	Failed.screening
C 6	0,05	1	Failed.screening

Tabella 6: Risultati dell'analisi di sequenziamento del genoma mitocondriale dei 13 campioni in esame. Spiegazione nel testo.

Dalla valutazione dei risultati ottenuti dal sequenziamento dei genomi mitocondriali ed in particolare, sulla base della copertura del DNA mitocondriale ottenuta per ciascun campione e sulla base dei risultati dell'analisi di autenticazione (*pattern* di *misincorporazione* e *pattern* di contaminazione) in collaborazione con i colleghi Americani abbiamo deciso di promuovere per la cattura ed il sequenziamento degli *SNPs* del genoma nucleare i 5 campioni: **C5**; **C34**; **C35**; **C33**; **C23**, come riportato in tabella 6.

4.2 Risultati dell'analisi del genoma nucleare

A partire dalle librerie *half-UDG* dei tredici campioni in esame inviate al dipartimento di Genetica dell'Università di Harvard, il gruppo del Prof. David Reich ha effettuato le analisi sul genoma nucleare. Dapprima è stata realizzata la determinazione genetica del sesso di tutti e 13 i campioni, poi è stata effettuata la cattura ed il sequenziamento di 1.240.000 *SNPs* del genoma nucleare soltanto per i campioni con risultati mitocondriali soddisfacenti, come descritto.

4.2.1 Determinazione Genetica di genere

Mediante un sequenziamento *shotgun* a bassa copertura è stata effettuata la determinazione di genere dei campioni, secondo il protocollo presente in letteratura [Skoglund P. 2013] [Fu Q. 2016] [Mittnik A, 2016]. L'assegnazione del genere, maschio oppure femmina, viene realizzata valutando la presenza o l'assenza del cromosoma Y rispettivamente. In particolare, a seguito del sequenziamento *shotgun* a bassa copertura, vengono selezionate le *reads* che mostrano la transizione C->T e che mappano sui cromosomi sessuali, andando a calcolare per ciascun campione, il valore del parametro R_y inteso come frazione del numero di allineamenti al cromosoma Y (n_y), diviso il numero totale di allineamenti ad entrambi i cromosomi sessuali (n_x+n_y). $R_y = n_y/(n_x+n_y)$. In figura 38 è riportata la trasposizione grafica degli allineamenti al cromosoma Y rispetto al numero totale di allineamenti ai cromosomi X ed Y per i 13 campioni in esame e l'indicazione dei generi assegnati.

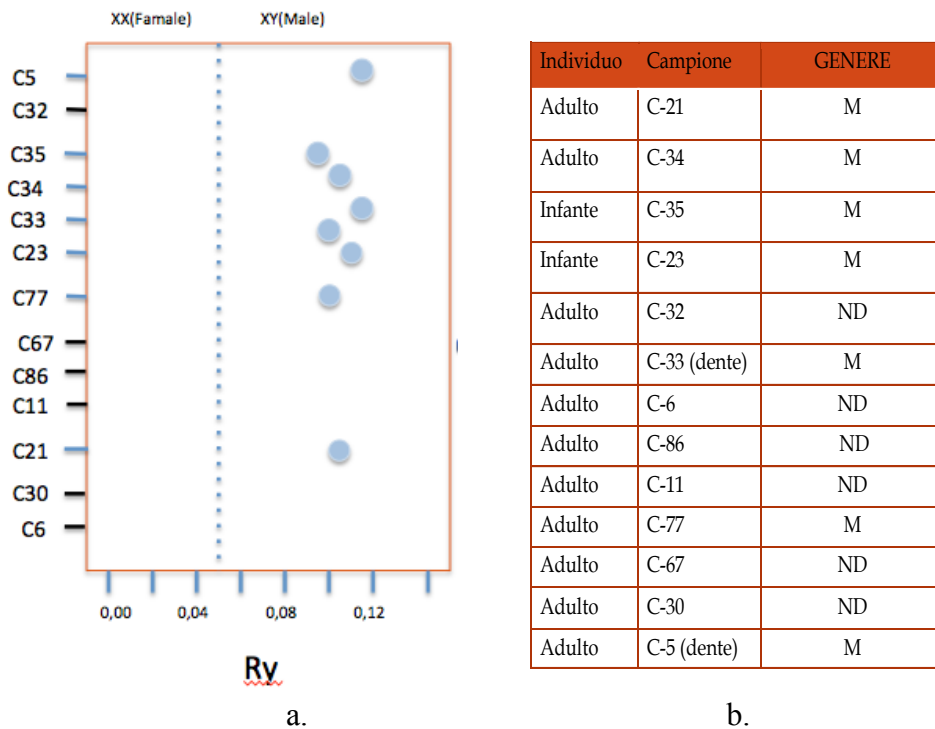


Figura 38: a. Frazione degli allineamenti al cromosoma Y. b. Genere identificato per ciascun campione. M (maschio) ND (non determinato)

Per alcuni campioni non è stato possibile determinare il genere a causa della bassa qualità del mappaggio.

4.2.2 Risultati dell'analisi degli *SNPs* nucleari

I cinque campioni C5; C34; C35; C33; C23 per i quali l'analisi del mtDNA ha dato risultati soddisfacenti in termini di copertura ed autenticità, sono stati processati per la cattura di **1.240.000 *SNPs*** del DNA nucleare mediante il protocollo, *Solution-phase Hybrid Selection, SHS*, [Gasc C. 2016] utilizzando le sonde appositamente disegnate per ibridare questi polimorfismi in molecole antiche [Qiaomei Fu et al. 2015; Qiaomei Fu. Meyer M., et al. 2013].

L'analisi di questa notevole quantità di dati di sequenza è tutt'ora in corso, ma sono stati ottenuti già i primi risultati. In tabella 7 sono riportati i 5 campioni esaminati con indicazione della copertura media delle regioni nucleari d'interesse ed il numero di polimorfismi identificati. Come atteso a copertura maggiore corrisponde un numero di *SNPs* identificati maggiore.

Campione	Sito di recupero	genere	mtaplo	copertura	n° <i>SNPs</i>
C33	Casa bracciale d'Oro	M	N1b1a	0.09	93727
C34	Casa bracciale d'Oro	M	U1a	0.437	364533
C23	Casa bracciale d'Oro	M	H	0.309	286023
C35	Casa bracciale d'Oro	M	T2c	0.054	62030
C5	Villa dei Misteri	M	H	0.047	53739

Tabella 7: campioni esaminati per gli *SNPs* nucleari.

Il numero di polimorfismi *SNPs*, identificati e condivisi tra i campioni in esame è stato utilizzato per realizzare un'analisi delle componenti principali, *PCA*, (*Principal Component Analysis*) dei dati genetici. Pertanto, applicando il metodo statistico, è stata realizzata una rappresentazione nello spazio bidimensionale delle frequenze alleliche negli individui di Pompei in relazione ad una banca dati di circa 900 individui moderni ed oltre 200 individui antichi da Europa e Asia Occidentale (*Western Eurasia*) presenti in letteratura, figura 39.

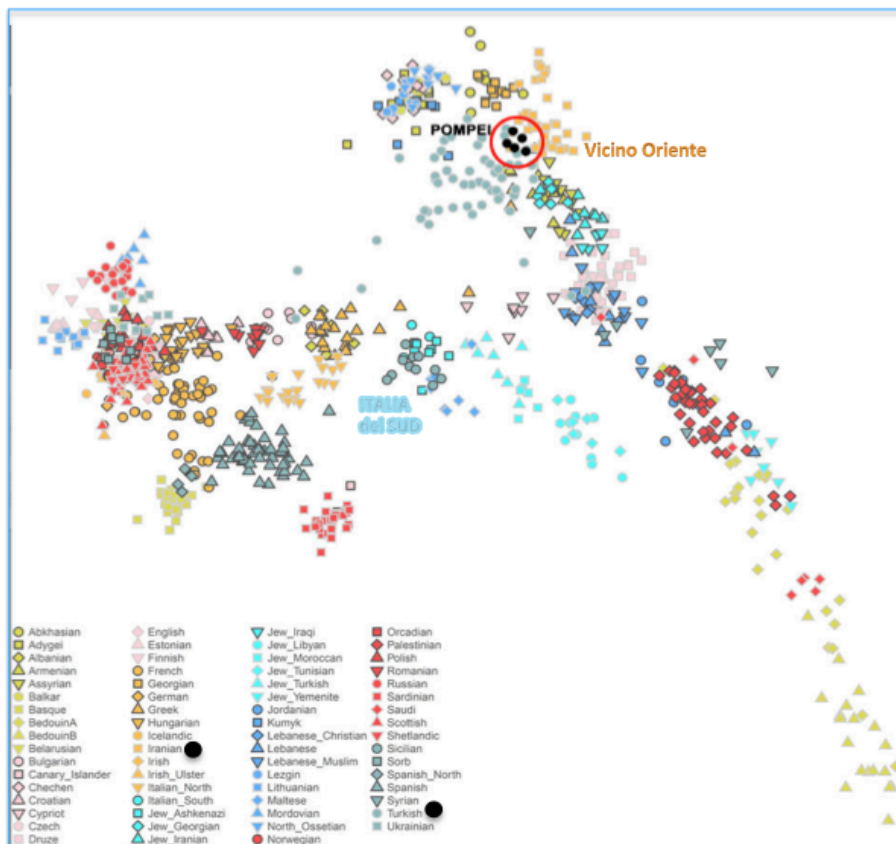
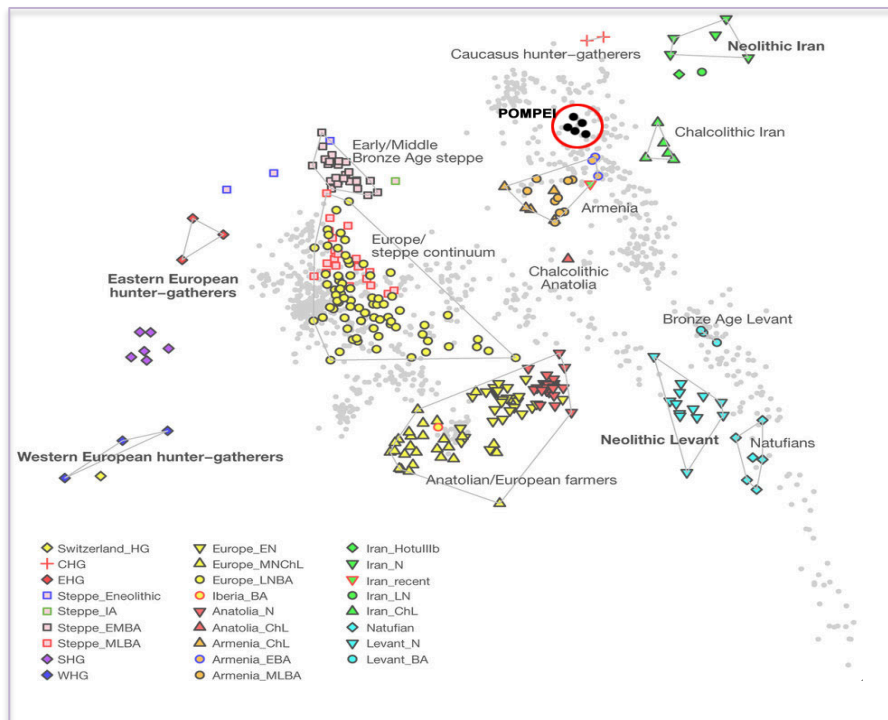


Figura 39: PCA dei dati genetici nucleari per i 5 campioni in esame. **In alto:** Frequenze alleliche degli SNPs negli individui di Pompeii in relazione ad un *database* di oltre 200

individui antichi (in evidenza e colorati) e 900 moderni (in grigio sullo sfondo). **In basso:** Frequenze alleliche degli *SNPs* negli individui di Pompei in relazione al *database* dei soli 900 individui moderni presenti anche nello schema in alto. Gli individui che hanno frequenze alleliche simili risultano più vicini nel grafico.

Dall'osservazione delle *PCA* si nota che a differenza di quanto atteso, le frequenze alleliche degli *SNPs* negli individui di Pompei in esame, non sono simili a quelle riscontrate negli Italiani del Sud, ma a quelle di individui del Vicino Oriente.

Pertanto, tale risultato ha posto la questione relativa all'origine geografica di questi individui. Se fossero persone nate nel vicino Oriente e giunte a Pompei a seguito degli scambi di uomini e merci, o se fossero già individui di seconda o terza generazione sul territorio. Per indirizzare quest'interrogativo abbiamo stretto una collaborazione con il gruppo del Prof. Douglas Kennett della Penn University, il quale mediante il sistema AMS, Spettrometria di Massa con Acceleratore, ha effettuato un'analisi isotopica sullo smalto dentale del campione **C5**. In particolare, ha analizzato i rapporti degli isotopi dello Stronzio e del Piombo ($^{87}\text{Sr}/^{86}\text{Sr}$; $^{208}\text{Pb}/^{204}\text{Pb}$; $^{207}\text{Pb}/^{204}\text{Pb}$; $^{206}\text{Pb}/^{204}\text{Pb}$).

Lo Stronzio è un metallo terrestre che si comporta in maniera simile al Calcio, (Ca^{2+}) nella maggior parte dei processi biologici ed ambientali. Come quest'ultimo viene assorbito dall'organismo umano attraverso l'acqua da bere ed il cibo e si accumula a livello dello scheletro e dei denti soprattutto durante l'accrescimento ed il rimodellamento di queste strutture [Killgrove K, Montgomery J., 2016]. Il Piombo, invece, è un metallo tossico che può essere assorbito sia per inalazione che per ingestione dall'ambiente circostante. Tracce di piombo possono essere riscontrate soprattutto a livello dei denti che ne incorporano quantità variabili durante la loro formazione. I livelli di Pb, nelle diverse aree geografiche del pianeta variano molto anche in base all'uso che viene fatto di questo metallo, e ciò si riflette sulla quantità del metallo riscontrabile a livello dentale [Kamenov D.George, 2014]. Tale caratteristica è stata sfruttata per seguire le migrazioni di soggetti umani di epoche passate [Sharpe AE, et al., 2016] [Shaw H., et al., 2016]. I rapporti isotopici riscontrati nello smalto dentale del campione C5 sono riportati in tabella 8.

Sample	Tooth	weight (g)	$^{87}\text{Sr}/^{86}\text{Sr}$	Sr error	$^{208}\text{Pb}/^{204}\text{Pb}$	error	$^{207}\text{Pb}/^{204}\text{Pb}$	error	$^{206}\text{Pb}/^{204}\text{Pb}$	error	voltage
C-5 (dente)	enamel	0,0535			38,5113	0,0016	15,6352	0,00062	18,4287	0,0007	5,9

Tabella 8: Rapporti isotopici identificati per il reperto dentale C5.

Dal confronto con i rapporti isotopici del Pb presenti in letteratura per campioni del centro-sud Italia e del Vicino Oriente, [Rich S., 2014] [Killgrove K. 2016] è stato osservato che i risultati ottenuti per il campione C5 si collocano più vicino a quelli dell'area del Vicino Oriente.

4.3 Datazioni al radiocarbonio

In collaborazione con il gruppo del Prof. Kennett della Penn University, è stata eseguita anche la datazione al ^{14}C dei campioni C5, C33, C34. I risultati sono riportati in tabella 9 e di seguito riporto brevi cenni sulla datazione al ^{14}C .

La tecnica del radiocarbonio (^{14}C o carbonio-14) permette di datare qualsiasi materiale di origine organica, derivante da qualcosa che sia stato vivo, (ossa, legno, stoffa, carta, semi, polline, pergamena e pellame in genere, tessuti e fluidi biologici) risalendo all'epoca della morte dell'individuo da cui proviene il campione, purché non siano passati più di 60.000 anni. Dopo tale periodo, il ^{14}C residuo è talmente esiguo da non permettere misure attendibili neppure con le tecniche più sofisticate [Lazzati, 2017]. Un elemento della tavola periodica si distingue da un altro per il numero di protoni del suo nucleo (che è identico in tutti gli atomi di quel dato elemento). Così tutti gli atomi di idrogeno hanno un protone nel nucleo, tutti quelli di carbonio ne hanno sei, tutti quelli di azoto ne hanno sette, e così via. Per uno stesso elemento, il numero di neutroni, invece, può variare da atomo ad atomo,; varia così il numero totale di particelle del nucleo, definito come numero di massa, mentre ovviamente resta immutato il numero di protoni (numero atomico) che identifica ogni elemento. Gli atomi di uno **stesso elemento**, che differiscono tra loro esclusivamente per il numero di massa, si chiamano **isotopi**, in quanto occupano lo stesso “posto” nella classificazione degli elementi chimici. Per ogni elemento esiste un isotopo più abbondante, che comprende la maggior parte degli atomi, mentre gli altri sono rappresentati in piccole quantità. Per quanto riguarda il **carbonio**, esistono in natura tre isotopi: il 99% degli atomi di carbonio è costituito da ^{12}C (carbonio-12), circa l'1% da ^{13}C (carbonio-13) e solamente poco più di un atomo ogni mille miliardi è ^{14}C (in caso di campioni antichi la frazione di ^{14}C è ancora più piccola, a causa del decadimento radioattivo) [Lazzati 2017]. Alcuni isotopi, come il ^3_1H (trizio), il $^{14}_6\text{C}$ (carbonio-14), il $^{131}_{53}\text{I}$ (iodio-131) e tanti altri, sono instabili in quanto il loro nucleo prima o poi si disintegra emettendo radiazioni; per questo motivo sono detti radioisotopi o radionuclidi. Il $^{14}_6\text{C}$, detto anche “radiocarbonio” è quindi un isotopo

radioattivo. Esistono diverse modalità di decadimento radioattivo durante le quali vengono emesse, a seconda dei casi, diversi tipi di particelle e, a volte, anche onde elettromagnetiche (raggi gamma). Il ^{14}C decade secondo la modalità beta- e senza emissione di raggi gamma: un neutrone si trasforma in un protone ed un elettrone che viene espulso dal nucleo sotto forma di radiazione beta-. Il nucleo si trova così con un neutrone in meno ed un protone in più, cioè con 7 protoni e 7 neutroni. In seguito alla disintegrazione, il ^{14}C si trasforma perciò in ^{14}N stabile e non radioattivo, con l'emissione di una radiazione beta-.

Il metodo di datazione al radiocarbonio si basa sul lento decadimento del ^{14}C , che costituisce un ottimo “orologio” archeologico; tale metodo fu messo a punto e pubblicato tra il 1947 ed il 1949 da un gruppo di chimici dell'Università di Chicago diretti da Willard Libby, che per questo ebbe il premio Nobel nel 1960. Le prime datazioni radiocarboniche si ebbero perciò a partire dal 1950.

Nonostante col passare del tempo ci si sia resi conto che i presupposti su cui si basava il metodo di Libby erano veri solo in prima approssimazione, la Comunità Scientifica ha deciso di continuare ad eseguire le datazioni secondo tali assunti, ottenendo la così detta “**datazione radiocarbonica convenzionale**” (*CRA*). Questa *Conventional Radiocarbon Age*, oggi chiamata “*Uncalibrated radiocarbon Age*” è la datazione fornita direttamente dai laboratori e non ancora calibrata. Tale datazione viene poi sottoposta ad una **calibrazione**, al fine di ottenere la data “reale” di calendario, confrontando la datazione convenzionale con quelle ottenute da campioni di età nota [Lazzati, 2017]. La data calibrata, può discostarsi anche di alcuni secoli da quella convenzionale e va considerata come la miglior stima della data “vera” [Lazzati, 2017]. La datazione radiocarbonica convenzionale (*conventional radiocarbon age* o *uncalibrated radiocarbon age*), non calibrata, è normalmente espressa in anni **BP** (*Before Present*, calcolati a ritroso a partire dal 1950, anno da cui si è iniziati a datare con il metodo del radiocarbonio) ed è sempre pubblicata insieme a quella calibrata sia perché “confrontabile” con quelle ottenute agli inizi dell'applicazione del metodo (quando non si effettuava ancora la calibrazione), sia perché più vicina ai “dati grezzi”, in quanto non ancora manipolata statisticamente tramite il processo di calibrazione [Lazzati, 2017]. La **datazione calibrata** è invece normalmente espressa come data di calendario (*calendar age*), in anni **BCE** (*Before Current Era*) e **AD** (*Anno Domini*) o **CE** (*Current Era*), a seconda che si tratti di anni prima o dopo Cristo.

Campione	¹⁴ Ccal	¹⁴ C
C33	19 calBCE – 118calCE	1955±20 BP
C34	43 calBCE – 54 calBCE	1995±20 BP
C5	160 calBCE – 4 calCE	2050±20 BP

Tabella 9: Datazione al radiocarbonio dei tre campioni indicati di cui due sono reperti dentali (C33 e C5).

La datazione al ¹⁴C ha confermato che si tratta di campioni antichi appartenenti ad individui deceduti con un certo grado di approssimazione insito nella matodica, in data compatibile con l'epoca dell'eruzione (79 D.C.).

5. Discussione

Nel presente lavoro di Dottorato, sono stati presi in considerazione sedici frammenti ossei e due reperti dentali prelevati durante il restauro di calchi umani nell'area archeologica di Pompei (*Insula Occidentalis*). Lo scopo era verificare la presenza di DNA endogeno ancora analizzabile all'interno di reperti aventi una storia tafonomica così peculiare e realizzare un'estesa analisi dei genomi mitocondriale e nucleare al fine di caratterizzare individui residenti nella città di Pompei all'epoca della grande eruzione. I reperti scheletrici a nostra disposizione si presentavano in frammenti, in commistione al materiale del calco, di dimensioni molto ridotte. A causa del precario stato di conservazione del materiale scheletrico, da cinque frammenti ossei non è stato possibile ottenere la quantità di polvere necessaria all'estrazione del DNA. Pertanto, l'analisi genetica mitocondriale è stata eseguita sui restanti tredici campioni. A seguito della cattura del DNA mitocondriale a mezzo di sonde oligonucleotidiche complementari, è stata ricostruita la sequenza del mtDNA per ciascun campione, mediante l'utilizzo di strumenti bioinformatici apposti per il DNA antico e mappaggio delle *reads* ottenute dal sequenziamento *NGS* con la sequenza di riferimento del mtDNA, la *rCRS*. Su tredici campioni esaminati, per nove è stato possibile ottenere una copertura mitocondriale sufficiente alla determinazione dell'aplogruppo (*software Haplofind*) ed alla valutazione dell'autenticità del profilo mitocondriale ottenuto (*software Mapdamage* e *ContamMix*). Sulla base dell'alta copertura mitocondriale e dell'esito positivo della valutazione di autenticità, i campioni C5; C34; C35; C33; C23 sono stati promossi per l'analisi degli *SNPs* del genoma nucleare realizzata presso il laboratorio del Prof. Reich. Qui, è stato realizzato anche un sequenziamento *shotgun* a bassa copertura sui tredici campioni analizzati per il DNA mitocondriale, ed è stata effettuata la stima delle *reads* mappanti sull'eterosoma Y al fine della determinazione di genere. Gli individui in esame sono risultati tutti di sesso maschile e quelli per cui si era ottenuta una bassa copertura mitocondriale, hanno prodotto anche in quest'analisi una qualità di mappaggio agli eterosomi troppo bassa per poter determinare il genere. Pertanto, dei campioni C32; C67; C86; C11; C30; C6 il genere rimane ignoto. Relativamente ai campioni C21, C23, C34, C35; questi sono stati prelevati da individui rinvenuti nel medesimo ambiente della Casa del bracciale d'Oro e gli Archeologi avevano supposto si trattasse di un nucleo familiare composto da padre, madre (recava in grembo uno dei due bambini ed al polso il bracciale che ha dato il nome alla casa) e due figli (C23 e C35 frammenti ossei da infanti). Le analisi genetiche hanno smentito l'ipotesi, in quanto la presunta madre è risultata un individuo di genere maschile ed i due bambini non sono tra loro imparentati per linea materna presentando un diverso aplogruppo mitocondriale.

Relativamente ai campioni C32 e C33, prelevati da individui rinvenuti abbracciati nel medesimo ambiente, gli Archeologi avevano supposto si trattasse di due sorelle. Anche in questo caso le analisi genetiche contraddicono l'ipotesi, in quanto uno dei due individui (campione C33) è risultato di genere maschile ed inoltre non condividono l'aplogruppo mitocondriale, pertanto non possono essere imparentati per linea materna.

Le analisi genetiche relative alla presunta famiglia ed alle presunte sorelle sono state realizzate anche dalla mia collega Elena Pilli presso il Reparto Investigazioni Scientifiche dei Carabinieri di Roma, la quale attraverso metodiche di pertinenza forense è giunta a risultati sovrapponibili.

I campioni C5, C34, C35, C33, C23, per i quali è stato possibile ottenere i più soddisfacenti risultati dall'analisi del genoma mitocondriale, sia in termini di copertura che di autenticità delle sequenze antiche ottenute, sono stati promossi alla cattura di 1.240.000 polimorfismi a singolo nucleotide del DNA nucleare. Allo scopo, sono state utilizzate sonde oligonucleotidiche appositamente disegnate dal gruppo del Prof. Reich, per ibridare questi polimorfismi in molecole di DNA con le caratteristiche transizioni

C->T al 5' terminale. Le analisi delle *reads* prodotte dal sequenziamento degli *SNPs* catturati sono tutt'ora in corso, ma già sono stati ottenuti i primi risultati. Sulla base delle frequenze degli alleli identificati e condivisi nei cinque Pompeiani è stato possibile effettuare un'analisi delle componenti principali (PCA) dei dati genetici. In particolare, dalla rappresentazione nello spazio bidimensionale delle frequenze alleliche negli individui di Pompei in relazione ad una banca dati di circa 900 individui moderni ed oltre 200 individui antichi da Europa ed Asia Occidentale, è emerso che i Pompeiani in esame hanno frequenze alleliche simili a quelle degli individui del Vicino Oriente. Questo risultato ha ingenerato l'ulteriore investigazione circa l'origine geografica dei nostri Pompeiani: è plausibile che fossero individui nati nel Vicino Oriente ed arrivati a Pompei in seguito agli scambi di uomini e merci? Oppure è più probabile che fossero già individui della seconda o terza generazione sul territorio? Per investigare tali quesiti è stata realizzata l'analisi dei rapporti degli isotopi dello Stronzio e del Piombo sullo smalto del reperto dentale C5, ad opera del gruppo del Prof. Kennett (Università della Pennsylvania). Gli isotopi di questi metalli possono accumularsi a livello dei denti durante la loro formazione, in percentuale variabile proporzionalmente all'entità di esposizione che è funzione della loro presenza nell'area geografica d'interesse. Il confronto dei rapporti degli isotopi del piombo trovati nel campione C5 rispetto a quelli presenti in letteratura relativi alla regione centro meridionale dell'Italia e del Vicino Oriente, rispettivamente, ha mostrato valori più simili a quelli della regione del *Near East*. Pertanto, è plausibile supporre che si trattasse di individui nati nella zona Orientale trasferiti poi a Pompei, ove presumibilmente si trovavano in veste di schiavi. Questo, in

concomitanza con i luoghi in cui sono stati ritrovati (l'individuo del campione C5 nella Villa dei Misteri e gli altri quattro nella Casa del Bracciale d'Oro) è anche compatibile con il dato storico. All'epoca dell'eruzione del 79 d.C. la città era ancora in ristrutturazione dopo il violento terremoto che l'aveva colpita diciassette anni prima ed a seguito del quale molti esponenti del patriziato erano andati via temendo per la propria incolumità, lasciando gli schiavi a ristrutturare le Domus gravemente danneggiate.

Mediante il sistema di spettrometria di massa, AMS, poi, il gruppo del Prof. Kennett ha realizzato la datazione al ^{14}C dei campioni C33, C5 e C34 ed i risultati ottenuti confermano che si tratta di reperti antichi di un periodo compatibile con quello dell'eruzione.

6. Bibliografia

Anderson S., Bankier A. T., Barrell B.G., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457 - 465.

Andrews RM., Kubacka I., Chinnery P.F., et al.. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23(2): 147.

Angela A. (2016). I tre giorni di Pompei. Rizzoli editore.

Bouwman S. A., Chilvers E.R., Brown A.K. (2006). Brief Communication: Identification of the Authentic Ancient DNA Sequence in a Human Bone Contaminated with Modern DNA. *Am. J. of Physical Anthropology* 131:428-431.

Budowle B. and van Daal A. (2008). Forensically relevant SNP classes. Review. *BioTechniques* 44:603-610.

Butler John M. (2009). Fundamental of forensic dna typing. Elsevier.

Butler John M. (2011). Advanced Topics in forensic DNA typing: methodology. Elsevier.

Briggs AW et al. (2007). Patterns of damage in genomic DNA sequences from a Neanderthal. *Proc.Natl Acad.Sci. USA* 104, 14 616–14 621.

Brotherton P., Endicott P, Sanchez Juan J., Beaumont M. (2007). Novel high-resolution characterization of ancientDNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research*, Vol. 35, No. 17 5717–5728.

Cann Rebecca, Stoneking M. Allan C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31- 36.

Caramelli D. Antropologia molecolare (2009). Manuale di base. Firenze University press.

Caramelli D., Lalueza-Fox C., Vernesi C., Lari M., Casoli A., et al. (2003). Evidence of Genetic dis- continuity between Neanderthal and 24,000-year-old anatomically moderns Europeans. *PNAS* 100 (11): 6593-7.

- Ciardello R.** (2012). Alcune riflessioni sulla Casa del Bracciale d'Oro a Pompei.
- Cipollaro M** , Di Bernardo G, Galano G. et al. (1998). Ancient DNA in human bone remains from Pompeii archaeological site. *Biochem Biophys Res Commun.* Jun 29; 247(3):901-4.
- Cock Peter J.**, Christopher J. Fields, Naohisa Goto, et al. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, Vol. 38, No. 6, pp. 1767–1771 .
- Cooper A.** and Poinar HN. (2000). Ancient DNA: Do It Right or Not at All. *Science.* Vol. 289, Issue 5482, pp. 1139.
- Dabney Jesse**, Knappb M., Glocke I. et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS* 15758–15763.
- Danecek P.**, Auton A., Abecasis Goncalo, et. al. (2011). The variant call format and VCFtools. *Bioinformatics*, 27 (15):2156-2158.
- De Franciscis A.** (1988). La casa di Caius Iulius Polibius. *Rivista di Studi Pompeiani* II,15-36.
- De Vos Arnold.**, De Vos Mariette, (1982). Pompei, Ercolano, Stabia, Roma. Editore Laterza e figli.
- Di Bernardo G**, Del Gaudio S., Galderisi U., (2009). Ancient DNA and Family Relationships in a Pompeian House. *Annals of Human Genetics* 73,429-437.
- Etienne Robert.** (1992). La vita quotidiana a Pompei, Arnoldo Mondadori Editore.
- Faerman M**, Kahila G, Smith P. et al. (1997). DNA analysis reveals the sex of infanticide victims. *Nature* 385(6613):212-3.
- Fu Q.**, Alissa Mittnik, Philip L.F. et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol.* 23(7): 553–559.
- Fu Q.**, Hajdinjak M., Moldovan O. T., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, pp.216-219.
- Fu Q.** Meyer M., Gao X., et al. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* 110, 2223–2227.

Fu Q., Posth C., Hajdinjak M., Martin Petr, et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200–205.

Gasc C., Peyretailade E and Peyret Pierre. (2016) Survey And Summary. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Research*, Vol. 44, No. 10 pp. 4504–4518 .

Gill P, Ivanov PL, Kimpton C. et al. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nat Genet.* 6(2):130-5.

Gill-King H. (1996). Forensic Taphonomy. Chapter 6. Chemical and Ultrastructural Aspects of Decomposition. Edito da William D. Haglund and Marcella H. Sorg. CRC press.

Ginolhac A, Rasmussen M, Gilbert M. TP., Willerslev E, Orlando L (2011). MapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics.* 27(15), 2153-5.

Gnirke A, Melnikov A, Maguire J. (2009). Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. *Nat Biotechnol.* 27(2): 182–189. doi:10.1038/nbt.1523.

Green R., Malaspinas AS, Krause J, Briggs AW, Johnson PL, Uhler C, et al. (2008). A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3):416-26.

Grimaldi Mario. Nuove decorazioni parietali dalla casa di Marco Fabio Rufo e dalla villa imperiale (2010). *The Journal of FastiOnline Documents and Research.*

Hagelber E., Clegg J.B. (1993). Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proc Biol Sci.* 252(1334):163-70.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC, (1984). DNA sequences from the quagga, an extinct member of the horse family, *Nature*, vol. 312, n° 5991, pp. 282–4.

Hofreiter M., David S., Hendrik N. Poinar. Ancient DNA. (2001). *Nature Reviews Genetics* 2, 353-359.

- Illumina** (2011). CASAVA v1.8.2 User Guide, (<https://support.illumina.com>).
- Jonsson H.**, Ginolhac A., et al, (2013). MapDamage 2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29 (13) pp.1682-1684.
- Kamenov D.**George, Brian L.Gulsonb (2014).The Pb isotopic record of historical to modern human lead exposure. *Science of The Total Environment* 490, pp. 861-870.
- Killgrove K**, Montgomery J (2016). All Roads Lead to Rome: Exploring Human Migration to the Eternal City through Biochemistry of Skeletons from Two Imperial-Era Cemeteries (1st-3rd c AD). *PLoS ONE* 11(2): e0147585.
- Kircher Martin** (2012). Analysis of High-Throughput Ancient DNA Sequencing Data. Part of the Methods in Molecular Biology book, ed. Shapiro B. Springer Science e Business Media, LLC.
- Krause J.**, Lalueza-Fox C., et al. (2007). The Derived FOXP2 Variant of Modern Humans Was Shared with Neanderthals. *Current Biology* 17, 1908–1912.
- Krause J**, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, Pääbo S (2010). A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Current Biology* 20, 231–236.
- Krings Matthias**, Anne Stone, Ralf W Schmitz et al. (1997). Neanderthal DNA Sequences and the Origin of Modern Humans. *Cell* 90, 19–30.
- Lazzati Marco**. *Il carbonio-14 e le datazioni archeologiche*. Aprile 2017.
- Li H.**, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-60.
- Li H.** and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics* 26(5): 589–595.
- Li R.** Forensic Biology 2nd edition. CRC press 2015.
- Lutz S** , Wittig H Weisser H.-J et al. (2000). Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Science International* 113, 97-101.
- Margulies M**, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J,

et al. (2005). Genome sequencing in microfabricated highdensity picolitre reactors. *Nature* 437, 376-380.

Maricic T, Whitten M, Paabo S (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoSOne* 5, 11.

Mascii M., C.R. Romeo (2012). Vincent Van Gogh. L'ultimo dipinto. Il fienile Protestante. Edizione Tecnostampa-Recanati.

Metzker Michael L. (2010). Sequencing technologies — the next generation. *Nature Reviews, Genetics* 11.

Meyer M, Kircher M (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc.* 10, 1101.

Mittnik A, Wang C-C, Svoboda J, Krause J (2016). A Molecular Approach to the Sexing of the Triple Burial at the Upper Paleolithic Site of Dolní Věstonice. *PLoS ONE* 11(10).

Nesheva DV (2014). Aspects of Ancient Mitochondrial DNA Analysis in Different Populations for Understanding Human Evolution. *Balkan J Med Genet.* 17(1): 5–14.

Okonechnikov K., Conesa A., Garcia-Alcalde F. (2016). QualiMap 2: advanced multisample quality control for high-throughput sequencing data. *Bioinformatics* 32(2): 292-4.

Pääbo S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 24; 314(6012):644-5.

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

Pilli E; Virgili A; D'errico G; Vespi G; Barni F; D. Caramelli; Lago G; Berti A (2006). Ottimizzazione di una nested PCR in campioni Low Copy Number. In: XVI CONGRESSO DEGLI ANTROPOLOGI ITALIANI, GENOVA, 29-31 OTTOBRE 2005, vol. 1, pp. 771-780.

Pliny the younger Letters:

(<http://www.perseus.tufts.edu/hopper/text?doc=Plin.+ep.+6.16>)

Rich S., Manning Sturt W.,P. Degryse, et al. (2016). Provenancing East Mediterranean cedar wood with the ⁸⁷Sr/⁸⁶Sr strontium isotope ratio. *Archaeol Anthropol Sci* 8:467–476.

Richards M. B., Macaulay V. A., et al. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* 62, 241–260.

Rohland N., Harney E., Mallick S., et al. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B* 370: 20130624.

Rohland N., Hofreiter M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols* 2(7), pp. 1756-1762.

Røyrvik E.C, Burgstaller J.P., Johnston I.G. (2016). mtDNA diversity in human populations highlights the merit of haplotype matching in gene therapies. *Mol Hum Reprod* 22 (11): 809-817.

Russell P. J., *Genetica*. EdiSES 2006.

Sancar A., Lindsey-Boltz LA., et al. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem.*

Sato M, Kuroiwa T. (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res.* 196(1):137-40.

Sharpe AE, Kamenov GD, Gilli A, Hodell DA, Emery KF, Brenner M, et al. (2016). Lead (Pb) Isotope Baselines for Studies of Ancient Human Migration and Trade in the Maya Region. *PLoS ONE* 11(11): e0164871.

Shaw H. , Montgomery J., Redfern R., et al. (2016). Identifying migrants in Roman London using lead and strontium stable isotopes. *J. of Archaeological Science* 6657, e68.

Skoglund Pontus, Storå Jan, Götherström Anders et al. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. of Archaeological Science* 40, 4477-4482.

Stoneking Mark, Jennifer J. Fontius, et al. (1997). *Alu* Insertion Polymorphisms and Human Evolution: Evidence for a Larger Population Size in Africa. *Genome Res.* 7(11): 1061–1071.

Sutovsky P, Moreno RD, Ramalho-Santos J, et al. (2000). Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod.* 63(2):582-90.

Torrioni A., Marie T. Lott, Margaret F. Cabell, et al. (1994). mtDNA and the Origins of Caucasians: identification of Ancient Caucasian-specific Haplogroups, one of which is Prone to a Recurrent Somatic Duplication in the D-Loop Region. *Am. J. Hum. Genet.* 55:760-776

Vai S., et al. (2015). DNA Sequencing in Cultural Heritage. *Top Curr Chem (Z) Review.*

Vernesi C., Caramelli D, Dupanloup I, et al. (2004). The Etruscans: A Population-Genetic Study. *Am J Hum Genet.* 74(4): 694–704.

Vianello D., Sevini F., Castellani G., et al. (2013). HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum Mutat.* 34 (9):1189-1194.

Vigilant L., Stoneking M, Harpending H, Hawkes K, Wilson AC. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253(5027):1503-7.

Wallace D.C. (1994). Mitochondrial DNA mutations in diseases of energy metabolism. *J Bioenerg Biomembr.* 26(3): 241-50.

Wallace DC., Brown MD, Lott MT (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238(1):211-30.

Wallace DC., Ye JH, Neckelmann SN, Singh G, et al. (1987). Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet.* 12(2):81-90.

Weyrich Laura S., Sebastian Duchene, Julien Soubrier et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 544, 357–361.

Wiegand P., Kleiber M. (2001). Less is more--length reduction of STR amplicons using redesigned primers. *Int J Legal Med.* 114(4-5):285-7.

RINGRAZIAMENTI

Ringrazio tutto il gruppo di Antropologia Molecolare dell'Università di Firenze, in particolare Elena Pilli, Martina Lari ed il Prof. Caramelli senza i quali questo lavoro non sarebbe potuto venire alla luce. Sono grata al Caso che ci ha fatto incontrare nel non troppo lontano 2014.

Ringrazio i colleghi Americani per la comunione d'intenti e le proficue collaborazioni.

Un Grande Grazie alla mia famiglia acquisita e d'origine per la fiducia ed il sostegno!

Infine ringrazio il lettore per la pazienza e l'interessamento.