

**DOTTORATO DI RICERCA IN BIOLOGIA
(XXX CICLO)**

**Studio di campioni antichi/degradati: la
tecnologia NGS e classica applicata ai fini
identificativi**

Tesi di

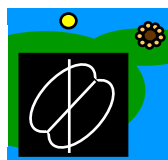
Elena Pilli



Coordinatore Prof. Alberto Ugolini

Tutor Prof. David Caramelli

(2017)





UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN BIOLOGIA
Curriculum Antropologia e Primatologia

CICLO XXX

COORDINATORE Prof. Alberto Ugolini

**Studio di campioni antichi/degradati: la
tecnologia NGS e classica applicata ai fini
identificativi**

Settore Scientifico Disciplinare BIO/08

Dottoranda

Dott.ssa Elena Pilli

Tutore

Prof. David Caramelli

Coordinatore

Prof. Alberto Ugolini

Anni 2014/2017

INDICE

INTRODUZIONE	7
1. I CAMPIONI ANTICHI/DEGRADATI	7
2. DIAGENESI DEL DNA ANTICO/DEGRADATO	10
3. PROBLEMATICHE CONNESSE ALL'ANALISI DEL DNA ANTICO/DEGRADATO	12
4. L'IDENTIFICAZIONE PERSONALE ED I POLIMORFISMI DEL DNA	15
5. I MARCATORI GENETICI IN AMBITO FORENSE	18
5.1 GLI <i>STRs</i>	19
5.2 GLI <i>SNPs</i>	21
5.3 IL DNA MITOCONDRIALE	22
6. L'ANALISI DEI MARCATORI FORENSI	25
6.1 TECNOLOGIA FORENSE CLASSICA	25
6.1.1 LA <i>PCR MULTIPLEX</i>	25
6.1.2 L'ELETTROFORESI CAPILLARE	27
6.1.3 IL SEQUENZIAMENTO SANGER	33
6.2 LA TECNOLOGIA <i>NGS</i>	35
6.2.1 LA TECNOLOGIA <i>NGS ILLUMINA</i>	37
6.2.2 LA TECNOLOGIA <i>NGS ION TORRENT</i>	39
6.2.3 METODI DI <i>TARGET ENRICHMENT</i> E SEQUENZIAMENTO <i>SHOTGUN</i>	43
<i>NEITHER FEMUR NOR TOOTH: PETROUS BONE FOR IDENTIFYING ARCHAEOLOGICAL BONE SAMPLES VIA FORENSIC APPROACH</i>	47
INTRODUZIONE	47
SCOPO DEL LAVORO	50
MATERIALI E METODI	51
1. I CAMPIONI	51
2. L'AMBIENTE DI LAVORO	52
3. PULIZIA E POLVERIZZAZIONE DEI CAMPIONI	53
4. ESTRAZIONE DEL DNA DA POLVERE D'OSSO	54
5. QUANTIFICAZIONE DEL DNA TRAMITE <i>REAL TIME PCR</i>	55

6. AMPLIFICAZIONE E TIPIZZAZIONE DEGLI <i>STRs</i>	57
7. IL SEQUENZIAMENTO	58
RISULTATI E DISCUSSIONE	60
CONCLUSIONI	69
BIBLIOGRAFIA	71

HUMAN IDENTIFICATION BY LICE: A NEXT GENERATION SEQUENCING CHALLENGE 77

INTRODUZIONE 77

SCOPO DEL LAVORO 83

MATERIALI E METODI 84

1. IL CAMPIONAMENTO	84
2. ISPEZIONE MICROSCOPICA	85
3. LE ANALISI GENETICHE	85
3.1 ESTRAZIONE DEL DNA	86
3.2 QUANTIFICAZIONE DEL DNA	87
3.3 <i>SNPs PROFILING</i> E PREPARAZIONE LIBRERIE	88
3.3.1 PCR EMULSIONE-CLONALE E SEQUENZIAMENTO	89
4. ANALISI BIOSTATISTICA	89

RISULTATI E DISCUSSIONE 93

CONCLUSIONI 103

BIBLIOGRAFIA 105

ANALISI DI DUE DENTI INVIATI IN BUSTA ANONIMA AD EDDA CIANO 109

INTRODUZIONE 109

SCOPO DEL LAVORO 114

MATERIALI E METODI 115

1. CAMPIONAMENTO	115
2. ISPEZIONE	116
3. TEST PRESUNTIVO E PRELIEVO PER ANALISI GENETICHE	116

4. PULIZIA E POLVERIZZAZIONE DEI CAMPIONI	117
5. ESTRAZIONE DEL DNA	117
6. QUANTIFICAZIONE DEL DNA	118
7. AMPLIFICAZIONE E TIPIZZAZIONE DEGLI <i>STRs</i>	118
8. PREPARAZIONE LIBRERIA <i>FORENSEQ</i>	119
RISULTATI E DISCUSSIONE	120
CONCLUSIONI	126
<i>NEXT GENERATION SEQUENCING E DNA CAPTURE: IL GENOMA MITOCONDRIALE DEI CALCHI DI POMPEI</i>	128
INTRODUZIONE	128
1. CENNI STORICI	128
1.1 L'ERUZIONE DEL 79 D.C.	130
1.2 GLI SCAVI ARCHEOLOGICI	133
SCOPO DEL LAVORO	138
MATERIALI E METODI	140
1. CAMPIONAMENTO	140
2. AMBIENTE DI LAVORO E PREVENZIONE DELLE CONTAMINAZIONI	145
3. PULIZIA E POLVERIZZAZIONE DEI CAMPIONI	145
4. ESTRAZIONE DEL DNA DA POLVERE D'OSSE	146
5. QUANTIFICAZIONE DEL DNA ESTRATTO	146
6. PREPARAZIONE DELLE LIBRERIE ILLUMINA	147
6.1 <i>BLUNT-END REPAIR</i>	148
6.2 <i>ADAPTER LIGATION</i>	148
6.3 <i>ADAPTER FILL-IN</i>	149
6.4 <i>INDEXING PCR</i> E <i>QPCR</i>	149
6.5 <i>ARRICCHIMENTO DELLA LIBRERIA TRAMITE PCR</i>	152
7. CATTURA DEL DNA MITOCONDRIALE	153
7.1 <i>PREPARAZIONE DELLE SONDE</i>	153
7.2 <i>INTERAZIONE DELLE SONDE CON LE BIGLIE MAGNETICHE</i>	154
7.3 <i>IBRIDAZIONE FRA LA LIBRERIA E LE SONDE</i>	154
7.4 <i>ELUIZIONE DELLA LIBRERIA CATTURATA</i>	155
7.5 <i>QUANTIFICAZIONE ED ARRICCHIMENTO DEL DNA CATTURATO</i>	156
8. ANALISI DEI DATI DI SEQUENZA	156

8.1 PACCHETTO FASTQPROCESSING/MIA	159
8.1.1 <i>ADAPTER TRIMMING E MERGING DELLE READS</i>	160
8.1.2 <i>MAPPING ASSEMBLY DELL'MTDNA</i>	161
8.4 ANALISI DEL <i>PATTERN</i> DI MISINCORPORAZIONE	162
8.5 ANALISI DELLE CONTAMINAZIONI	165
8.5.1 <i>SOFTWARE CONTAMMIX</i>	165
9. DETERMINAZIONE DELL'APLOGRUPPO MITOCONDRIALE	167
RISULTATI E DISCUSSIONE	169
CONCLUSIONI	196
BIBLIOGRAFIA	198
CONCLUSIONI FINALI	201
BIBLIOGRAFIA	204

RIASSUNTO

L'attività di ricerca svolta in questi tre anni di dottorato è stata condotta presso i Laboratori di Antropologia Molecolare dell'Università di Firenze in collaborazione con la Sezione di Biologia del Reparto Investigazioni Scientifiche dell'Arma dei Carabinieri di Roma. In questi tre anni mi sono occupata di svariati progetti centrati per lo più, fatta eccezione di alcuni progetti di filogenesi umana ed animale (dati non presenti in questa tesi ma che hanno permesso di realizzare pubblicazioni scientifiche), sull'analisi di campioni antichi/degradati al fine di identificarli e di investigare le eventuali relazioni di parentela esistenti fra gruppi di individui, coniugando così le competenze acquisite sui campioni antichi/degradati con quelle proprie dell'ambito identificativo/forense. In particolare, ho investigato la possibilità di utilizzare la rocca petrosa di campioni antichi datati VI-VII sec d.C., quale substrato biologico per l'identificazione personale attraverso la tipizzazione degli *STRs* (*Short Tandem Repeats*) mediante metodica forense classica. Tale studio è stato effettuato anche al fine di valutare la possibilità di sfruttare tale metodica classica, più rapida ed economica rispetto alla tecnologia *NGS* (*Next Generation Sequencing*), nello studio delle relazioni di parentela fra individui antichi. Inoltre mi sono occupata di investigare le relazioni di parentela per linea materna (analisi DNA mitocondriale) esistenti fra due gruppi di individui rinvenuti a Pompei, i cui resti ossei sono stati prelevati dai calchi in gesso al momento del restauro degli stessi. Per tale tipologia di analisi è stata utilizzata la tecnologia *NGS* (Illumina) associata ad un protocollo, specifico per

campioni altamente degradati, di cattura mediante sonde specifiche del genoma mitocondriale di interesse. L'analisi dei dati è stata effettuata mediante *pipeline* specifiche per i campioni antichi. Mi sono inoltre dedicata a due *casework* di interesse forense. Nel primo caso è stato possibile, tramite tecnologia *NGS* (*Thermo Fisher*), identificare il DNA di scarsa qualità e quantità presente nel tratto gastrointestinale di insetti appartenenti alla sottospecie di *Pediculus humanus capitis*. Nel secondo *casework* di interesse storico è stato possibile analizzare due denti presumibilmente appartenuti a Mussolini al fine di ottenere un profilo *STR* da confrontare con il campione di confronto di un bis nipote di Mussolini. Infine, anche se non presente in questa tesi a causa del fatto che il lavoro è preliminare, ho effettuato il disegno di un *pool* di sonde utili per la cattura del cromosoma Y al fine di poter mettere a punto un protocollo idoneo per indagare le relazioni di parentela per linea paterna di campioni ossei antichi. Tali sonde verranno impiegate per esempio nel Progetto di Leonardo da Vinci per investigare la sua linea maschile.

INTRODUZIONE

1. I campioni antichi/degradati

Fino a qualche tempo fa il nostro passato evolutivo veniva investigato direttamente, attraverso lo studio dei fossili o, indirettamente, attraverso lo studio delle forme e dei geni di individui contemporanei cercando poi di ricostruire i fenomeni biologici che li avevano portati a differenziarsi. Non era quindi possibile avere informazioni genetiche dirette su animali, piante e uomini che avevano popolato la terra prima di noi. Da circa 20 anni però le cose sono cambiate. Nonostante tuttavia esistano ancora molte problematiche connesse con le analisi di campioni antichi associate con la natura biologica del campione stesso, i ricercatori sono riusciti nel tempo ad analizzare direttamente il DNA di individui morti anche migliaia di anni fa (*Caramelli et al. 2003; Lalueza-Fox et al. 2005; Caramelli et al. 2006; Lalueza-Fox et al. 2006; Orlando et al. 2006; Krings et al. 2000; Ovchinnikov et al. 2000; Schmitz et al. 2002*). Dopo la morte di un organismo, il DNA contenuto nelle sue cellule va incontro ad una serie di processi degradativi, causati dalla lisi dei legami tra i nucleotidi (le quattro basi che formano il DNA: Adenina, Timina, Citosina e Guanina), che comportano frammentazioni, alterazioni e delezioni irreparabili della molecola del DNA (*Gilbert et al. 2003 A; Gilbert et al. 2003 B*). Con il termine DNA antico (aDNA = ancient DNA) si indica il materiale genetico proveniente da resti di organismi morti o da parti di essi, anche se ancora non vi è pieno accordo fra gli studiosi sul lasso di tempo che deve intercorrere fra la morte

dell'organismo ed il prelievo del suo DNA affinché questo possa essere considerato antico. Sussiste però un sostanziale accordo nel considerare come antico il DNA che viene estratto da specie estinte: il DNA rinvenuto nei fossili o nei sub-fossili di specie estinte è indiscutibilmente da considerarsi antico.

Per quanto riguarda l'uomo però, la definizione di DNA antico è ancora controversa; infatti, dal momento che non ci si può riferire alla specie umana come ad una specie estinta, risulta ancora più complesso riuscire a dare un'accezione chiara all'aggettivo "antico". Alcuni ricercatori sostengono che la definizione di DNA antico dovrebbe tenere in considerazione il grado di deterioramento dell'organismo: solo quando questo si è alterato al punto tale da rendere necessaria l'analisi del DNA per la sua identificazione, allora, si potrebbe o dovrebbe parlare di DNA antico. Tale criterio tuttavia non tiene in considerazione il fatto che i tempi di deterioramento variano enormemente in base alle condizioni ambientali in cui il corpo si viene a trovare dopo la morte. Da quanto appena esposto appare chiaro come dare una definizione univoca e precisa al DNA antico non sia semplice e soprattutto possa rischiare di "invadere" il campo di altre discipline quale per esempio la genetica forense. Secondo tale definizione infatti una vittima di omicidio il cui corpo si sia rapidamente decomposto così da renderlo irriconoscibile potrebbe essere erroneamente considerato un caso di DNA antico. In realtà, al di là del grado di conservazione del corpo e degli "aggettivi" che possono essere dati al suo DNA, nel caso di una vittima di un omicidio non bisogna dimenticare di essere di fronte ad un caso

giudiziario che deve essere trattato da specialisti esperti nel settore con il rigore scientifico dovuto a tale tipo di indagine.

Intorno agli anni 80 del secolo scorso, nasce l'idea, da parte di un gruppo di ricercatori quali *Russell Higuchi*, *Svante Paabo* e *Mark Stoneking*, di provare ad estrarre materiale genetico da campioni archeologici e fossili. Ciò che all'inizio sembrava essere pura utopia, un semplice "gioco" intellettuale di alcuni appassionati ricercatori, ben presto divenne un importante campo, in crescente sviluppo, della moderna biologia molecolare, grazie al quale è oggi possibile studiare il patrimonio genetico di specie estinte e metterlo a confronto con quello delle specie attuali allo scopo di far luce sull'origine dell'umanità. Nasce così, all'interno dell'antropologia molecolare -disciplina che cerca di ricostruire la storia evolutiva della nostra specie attraverso l'analisi di molecole complesse, quali il DNA, di individui viventi della specie umana e delle specie a noi più affini, quali le grandi antropomorfe- un nuovo settore di ricerca che è quello del DNA antico. Ed è proprio grazie ai progressi relativamente recenti che si sono avuti nel campo delle biotecnologie ed anche alla messa a punto di sofisticate tecniche di analisi del materiale genetico, che questa nuova branca dell'antropologia molecolare ha potuto espandersi e contribuire significativamente alla ricostruzione della nostra storia evolutiva a partire dall'indagine diretta del DNA di individui appartenuti a specie e/o popolazioni ormai estinte o morti anche da migliaia di anni. Inoltre, recentemente, ciò che solo pochissimi anni fa veniva considerato impossibile da realizzare, ricostruire cioè l'intero genoma di un organismo, oggi, grazie alle

innovative tecniche messe a disposizione dalla scienza, è diventato possibile (Krause et al. 2010; Zeyland et al. 2013; Prüfer et al. 2017; Skoglund et al. 2017; Unterländer et al. 2017).

2. Diagenesi del DNA antico/degradato

Con il termine diagenesi si indicano in geologia i cambiamenti a cui va incontro un sedimento con il passare dei secoli fino alla sua completa trasformazione in roccia. L'applicazione dello stesso termine al DNA antico consente di riassumere l'insieme di tutti i cambiamenti chimici a cui va incontro la doppia elica e che portano alla sua degradazione/frammentazione. La doppia elica completamente idratata, pur essendo dotata di una elevata stabilità chimica e meccanica, una volta privata dei meccanismi di riparazione presenti in una cellula vivente, va incontro ad una demolizione spontanea che la riduce in piccoli frammenti, la cui velocità dipende dall'ambiente in cui il campione si viene a trovare al momento della sua morte.

La principale reazione di degradazione a cui è sottoposto il DNA è dovuta principalmente all'idrolisi dei legami tra lo zucchero ed il gruppo fosfato lungo i filamenti della doppia elica, che portano alla formazione di estremità a singola elica ed all'idrolisi dei legami N-glicosidici tra la base azotata e lo zucchero 2-deossiribosio che porta alla conseguente perdita della base azotata e creazione di un sito abasico. Poiché la velocità con cui le basi puriniche (guanina e adenina) vengono liberate è 20 volte superiore a quella delle pirimidine (citosina e timina), il processo di idrolisi delle basi viene solitamente indicato come

“depurinazione”. Una volta creato un sito abasico, in sua corrispondenza il filamento si indebolisce e può andare incontro a rottura in pochi giorni. In relativamente breve tempo questo meccanismo porta ad una frammentazione sempre più estesa della doppia elica. Ovviamente l'entità della degradazione può essere più o meno spinta a seconda del campione e, soprattutto, in rapporto agli stress ambientali (temperatura, pH, forza ionica del mezzo che, se molto alta, può favorire una riduzione da 5 a 10 volte della velocità del processo) a cui il campione è stato sottoposto nel corso del tempo. Ambienti in cui la temperatura è costantemente bassa favoriscono la conservazione del materiale genetico: non a caso, infatti, i campioni più antichi finora analizzati con successo provengono dal *permafrost* (Rasmussen *et al.* 2010). Insieme al congelamento anche l'essiccazione e l'alta concentrazione salina consentono una buona conservazione del materiale genetico. Ciò nonostante, calcoli compiuti sulla cinetica di degradazione indicano che frammenti delle dimensioni di 100-500 pb non sopravvivono per più di 10.000 anni in climi temperati e per un massimo di 100.000 anni a latitudini più fredde. Quindi, nonostante diversi studi continuino a riportare dati su cellule, soprattutto batteriche, antiche di diversi milioni di anni, rimane il dubbio se, anche in condizioni ideali, la molecola del DNA possa sopravvivere per più di 1 milione di anni.

Nonostante la depurinazione rappresenti il principale meccanismo di degradazione del DNA in soluzione, tra gli altri tipi di cambiamenti, attivi sugli organismi viventi, a cui è sottoposto il DNA, l'azione dell'ossigeno esercita un ruolo apprezzabile nella diagenesi del DNA

antico in resti sepolti. L'ossigeno infatti è responsabile della creazione di basi azotate modificate che possono portare, anche in questo caso, a frammentazione della molecola del DNA.

3. Problematiche connesse all'analisi del DNA antico/degradato

Il trattamento dei campioni antichi, l'analisi del DNA in essi contenuto come pure l'interpretazione dei risultati ottenuti sono caratterizzate da molteplici problematiche connesse principalmente con lo stato di estrema degradazione della doppia elica, il basso numero di molecole endogene presenti in tale tipo di campioni, la presenza di DNA contaminante di natura umana e non umana e la presenza di inibitori, specie chimiche di natura eterogenea, sia endogene che esogene, che possono interferire con le analisi inficiandone anche il risultato. Uno dei principali problemi quindi a cui lo studioso di DNA antico deve far fronte è, come già accennato, la degradazione della doppia elica in piccoli frammenti di dimensioni variabili. Un'altra criticità connessa con l'analisi del DNA antico è data dalla scarsa quantità di materiale genetico endogeno presente in tale tipo di campioni. Tale difficoltà di analisi è stata superata, inizialmente, grazie all'invenzione della reazione a catena della DNA polimerasi (PCR) in grado di riprodurre/moltiplicare in vitro frammenti di DNA. La PCR ha consentito di migliorare la sensibilità delle metodiche utilizzate per lo studio del DNA, permettendo di selezionare il target di interesse anche quando presente in scarsa quantità. Ma, se da un lato la PCR ha permesso un'analisi più immediata e precisa del DNA antico rendendo possibile la selezione del frammento di

interesse, dall'altro però ha fatto accrescere in modo esponenziale, data l'estrema sensibilità della metodica, le problematiche connesse con la contaminazione del campione. Nel corso della sua storia tafonomica un campione antico va inevitabilmente incontro a contaminazione che può essere dovuta sia a microrganismi che, crescendo in maniera saprofitica a spese del materiale cellulare, lo degradano e lo impoveriscono del DNA originale, arricchendolo, nel contempo, del loro, sia a manipolazioni del campione compiute al momento della deposizione dello stesso o nelle successive fasi di recupero e studio. Nel caso di resti umani, però, le tecniche di analisi molecolare hanno permesso di discriminare tra la componente primaria (endogena) degli acidi nucleici e quella secondaria esogena di tipo batterico o fungino. Il vero problema, allorché si voglia analizzare il DNA di un resto umano antico è pertanto rappresentato solo dalle manipolazioni compiute al momento della deposizione del campione e poi, in successione, dall'archeologo, dal personale del museo, dallo studioso di morfologia ed infine dall'antropologo molecolare. Per effetto di queste manipolazioni il materiale antico può subire un progressivo "inquinamento" da parte di materiale cellulare umano e quindi di DNA a seguito, per esempio, della desquamazione epidermica. Il DNA quindi che si estrae da reperti antichi può essere considerato una miscela di DNA endogeno, più o meno degradato, e di DNA contaminante. Le contaminazioni vengono tenute sotto controllo lavorando in condizioni di sterilità sia al momento del prelievo dei resti sia durante le fasi di laboratorio considerate più delicate. È bene infatti sottolineare come le contaminazioni originate da DNA moderno,

verosimilmente introdotte durante le attività sperimentali di laboratorio, possono essere controllate e minimizzate adottando tutta una serie di criteri particolarmente rigidi che sono stati pubblicati da *Cooper e Poinar* nel 2000 sulla rivista *Science* (*Cooper and Poinar, 2000*). Un'ulteriore difficoltà che si incontra nello studio di campioni antichi è data dalla presenza di molecole che possono, in vario modo, interferire con l'analisi del DNA antico. Molto spesso queste sostanze sono componenti dei fluidi biologici quali per es. emoglobina, collagene di tipo I, urea ed l'eparina, costituenti organici e fenolici di alcuni cibi, glicogeno, grassi, Ca^{2+} , composti ambientali quali gli acidi umici e fulvici, tannini, metalli pesanti o reagenti impiegati nelle fasi sperimentali come, un esempio fra tutti, i composti fenolici. La presenza di queste molecole durante le varie fasi sperimentali di analisi del DNA antico possono causare l'insorgenza di falsi-negativi, risultati cioè assenti anche in presenza di materiale genetico proprio del campione oggetto di studio. Nonostante in letteratura venga riportata un'ampia gamma di molecole definite inibitori in quanto in grado di impedire l'ottenimento di un risultato genetico apprezzabile, le modalità d'azione di molte di esse rimangono tuttora oscure.

Ad oggi tuttavia, il problema della frammentazione della doppia elica del DNA antico, del basso numero di molecole endogene e della presenza di DNA contaminante, batterico e umano, può essere superato grazie alla messa a punto di nuovi sistemi di indagine molecolare (NGS *Next Generation Sequencing*) in grado di valutare la qualità e di conseguenza la quantità del DNA endogeno non contaminato presente all'interno del

campione analizzato. In questo modo anche campioni che sono stati manipolati nel corso degli anni (esempio quelli delle collezioni museali o paleo-antropologiche) possono essere studiati senza il timore di incorrere nell'analisi delle contaminazioni eventualmente presenti.

4. L'identificazione personale ed i polimorfismi del DNA

L'identificazione personale è l'insieme di attività che vengono compiute su viventi, cadaveri, resti umani per attribuire l'identità ad una persona, o ai suoi resti o alle sue tracce biologiche rinvenute su una scena di un crimine. L'identificazione si effettua per esempio per i latitanti ed i criminali, per i casi di sospetta sostituzione di neonati, di rapimento, traffico e commercio di minori, soggetti affetti da amnesia, soggetti in stato di coma, tracce biologiche rinvenute sulla scena di un crimine, scambi di persone in ambito assicurativo ed altre situazioni che trovano riscontro a vario titolo nel codice penale. L'identificazione della persona nei cui confronti vengono svolte indagini giudiziarie è prevista dall'articolo 349 del Codice Penale. La legge 155/2005 ha apportato modifiche significative a questo articolo, dando facoltà alla polizia giudiziaria di procedere al prelievo di un campione biologico per l'esame del DNA o con il consenso dell'interessato, ovvero coattivamente previa autorizzazione del pubblico ministero.

Affinché si possa procedere all'attribuzione di una identità ad un vivente, ad un cadavere, a resti scheletrici o a tracce biologiche occorre disporre di elementi di raffronto, rappresentati, caso per caso, da fotografie, descrizioni, radiografie, schemi dentari, materiale biologico per

tipizzazioni genetiche, ecc., appartenuti alla persona nei confronti della quale si indirizza l'identificazione (dal latino *identicum facere*) o eventualmente ai suoi familiari.

Di fronte a resti scheletrici o a corpi degradati dalla putrefazione o dai traumi, le indagini mirano preliminarmente all'identificazione di caratteri generici (identificazione generica), classicamente indicati nell'età, sesso, statura, costituzione somatica, sviluppo corporeo e stimate professionali, al fine di costruire un profilo somatico che possa consentire di restringere il campo della ricerca. Tuttavia, in tutti quei casi in cui quanto appena descritto non può essere effettuato a causa per esempio di fenomeni putrefattivi o di alterazione dei lineamenti che modificano le fattezze fisionomiche del viso e dei tessuti rendendo impossibile l'identificazione visiva, è possibile ricorrere all'identificazione genetica tramite l'analisi del DNA e sua comparazione con campioni di confronto. L'identificazione genetica risulta inoltre necessaria in tutti quei casi in cui sulla scena di un crimine vengono rinvenute tracce biologiche utili per identificare il colpevole di un reato.

Intorno agli anni 80, cioè diversi anni dopo la scoperta della struttura del DNA, l'inglese *Alec Jeffrey* notò che alcune regioni del DNA contenevano sequenze altamente ripetute e notò anche che queste non si ripetevano allo stesso modo in ogni individuo ma che il numero di tali ripetizioni variava se si consideravano individui diversi. Si accorse così che esisteva una piccola parte del nostro DNA che ci rendeva unici. Tali ripetizioni erano note con il nome di *VNTR* (*variable number tandem repeats*). Per poterle studiare, *Alec Jeffreys*, utilizzò una tecnica

denominata *RFLP* (*restriction fragment length polymorphism*), che consisteva nell'impiego di particolari enzimi in grado di tagliare nelle regioni circostanti le VNTR. Queste ripetizioni risultavano avere un maggior potere di discriminazione rispetto alla determinazione del gruppo sanguigno che veniva effettuata in precedenza anche se, al pari del gruppo ABO, questo metodo necessitava di grandi quantità di materiale genetico di partenza, condizione difficile da avere quando si analizzano campioni rinvenuti sulla scena di un crimine. Questa tecnica aveva inoltre lo svantaggio di impiegare tempi lunghi per la sua completa realizzazione. Grazie alla scoperta del DNA *fingerprinting*, metodo di indagine genetica (detto anche impronta del DNA) che permette di identificare il DNA proveniente da individui diversi, sono stati risolti casi di violenza carnale e di omicidio (spesso arrivati alla cronaca per la loro efferata violenza): basti pensare che la prima applicazione di questa tecnica eseguita da *Sir Jeffrey* avvenne in un caso di stupro e omicidio di due giovani ragazze che si è poi concluso con l'arresto del colpevole, *Colin Pitchfork*, a *Narborough* (in Gran Bretagna) nel 1987. Il famoso caso è passato alla storia in quanto *Colin Pitchfork* è stato il primo criminale "incastrato" dal suo DNA dal momento che il campione di sperma di confronto corrispondeva perfettamente ai campioni prelevati dalle due ragazze morte. Questo fu un caso di identificazione particolarmente significativo anche perché, se non presente l'analisi genetica, sarebbe stato inevitabilmente condannato un innocente. Infatti l'analisi di *Jeffrey* non solo identificò il vero assassino, ma permise anche

di scagionare un innocente *Richard Buckland*, inizialmente sospettato degli omicidi.

Appurata l'affidabilità del DNA quale strumento d'indagine, sono stati compiuti notevoli sforzi da parte degli studiosi al fine di riuscire a trovare nuovi e migliori marcatori molecolari, per rendere più efficienti e maggiormente discriminanti le analisi. Il concetto di tipizzazione del DNA diede dunque inizio allo studio dei polimorfismi del DNA, ovvero variazioni nelle sequenze di DNA presenti nella popolazione con una frequenza maggiore dell'1%. Lo studio dei polimorfismi del DNA portò all'identificazione di regioni non codificanti costituite da sequenze altamente ripetute (polimorfismi di lunghezza) chiamate microsatelliti (*Short Tandem Repeats STRs*). In virtù della loro relativamente corta sequenza di ripetizione (da 2 a 5 paia di basi) tali regioni risultavano facilmente differenziabili sulla base del peso molecolare. L'analisi degli *STRs* unita alla tecnologia della *PCR* ha permesso di raggiungere ottimi livelli sia in termini di velocità di ottenimento dei risultati, sia in termini di sensibilità. Se con i metodi precedenti erano infatti necessarie cospicue quantità di materiale biologico di partenza, come ad esempio grandi macchie di sangue con DNA ben conservato, oggi, grazie all'analisi degli *STRs*, sono sufficienti piccole quantità di campione, come ad esempio poche cellule, per ottenere il profilo di DNA desiderato.

5. I marcatori genetici in ambito forense

I marcatori genetici utilizzati in ambito forense sono quelli del DNA nucleare -polimorfismi di lunghezza e polimorfismi di sequenza (Figura

1) dei cromosomi autosomici e dei cromosomi sessuali X e Y- ereditati, gli autosomici, per ciascun individuo per metà dal padre e per metà dalla madre in seguito a ricombinazione genetica, ed i marcatori del DNA mitocondriale, come quelli del cromosoma Y, in maniera uniparentale senza subire ricombinazione



Figura 1: I polimorfismi del DNA

5.1 Gli STRs

I polimorfismi di lunghezza denominati microsatelliti o *short tandem repeats* (STR) sono caratterizzati da unità di ripetizione molto corte di 2-5bp e presentano un alto grado di polimorfismo dovuto proprio alla variazione della lunghezza nel numero delle unità ripetute. Sulla base del numero di nucleotidi presenti nella ripetizione i più utilizzati in ambito forense sono gli STRs tetranucleotidi poiché molto discriminanti. Gli STRs rappresentano, all'attuale "stato dell'arte", il metodo d'elezione per l'identificazione personale attraverso marcatori biologici. Essi vengono utilizzati come marcatori in ambito forense in quanto (Tagliabracci 2010):

- presentano una buona variabilità consentendo l'identificazione senza ambiguità delle varianti e della loro distribuzione all'interno delle differenti popolazioni;
- hanno un meccanismo di trasmissione di tipo mendeliano;
- hanno ridotte dimensioni molecolari, offrendo pertanto la possibilità di tipizzare anche tracce biologiche degradate.

Oltre alla distinzione per lunghezza delle unità ripetute gli STRs possono differenziarsi anche per il tipo di microsatellite analizzato:

- ripetizioni semplici, costituite da unità ripetute identiche sia in lunghezza che in sequenza;
- ripetizioni composte, costituite da due o più ripetizioni semplici adiacenti;
- ripetizioni complesse cioè ripetizioni che possono contenere molti blocchi di ripetizioni costituiti da diverse unità ripetute, frapposte da sequenze variabili. Sono quindi costituiti da una regione ripetuta e dalle due regioni fiancheggianti uguali per tutti gli alleli dello stesso STR.

La nomenclatura utilizzata per identificare i diversi *loci STR* è specifica per ciascuno di essi. Se un marcatore fa parte o cade all'interno di un gene, è il gene stesso a dare il nome al microsatellite. Per esempio il marcatore TH01 è posizionato all'interno del gene *tyrosine hydroxylase* che mappa sul cromosoma 11. La sigla "01" nel suo nome deriva dalla regione, contenente il numero variabile di ripetizioni, situata all'interno

dell'introne 1 del gene che codifica per la tirosina idrossilasi. A volte il prefisso HUM è incluso nel nome del *locus* ad indicare che fa parte del genoma umano, quindi il marcatore TH01 potrebbe essere indicato come HUMTH01. Tutti i marcatori collocati in regioni extra-geniche vengono definiti con la loro posizione sul cromosoma. Ad esempio nel marcatore D5S818, marcatore che non si trova all'interno di un gene, la lettera D sta per DNA mentre il successivo numero 5 si riferisce al cromosoma 5, la lettera S indica che il marcatore di DNA è a singola copia, mentre il numero finale indica la posizione del marcatore sul cromosoma 5.

5.2 Gli SNPs

La maggior parte dei polimorfismi del DNA non è dovuta però solo agli *STRs*, bensì ad un'altra fonte di variabilità presente abbondantemente nel genoma umano, i polimorfismi di sequenza, ovvero gli *SNPs* (*Single Nucleotide Polimorphism*). Tali marcatori costituiscono una variazione puntiforme del DNA poiché presentano la variazione di una singola base, quale una sostituzione, un'inserzione o una delezione ad una singola posizione nel genoma, per cui due alleli si possono assortire in 3 differenti genotipi. Attualmente sono noti circa 3 milioni di *SNPs* distribuiti uniformemente su tutto il genoma ma si calcola ve ne siano circa 10 milioni. Essendo la principale fonte di variazione umana, attualmente, si sta iniziando ad utilizzare gli *SNPs* per la realizzazione di nuovi sistemi di identificazione personale. L'utilità nel campo delle indagini forensi è data dal fatto che, rispetto agli *STRs* che richiedono amplificati di dimensioni maggiori di 100 bp, gli *SNPs* necessitano, per

essere tipizzati, di amplificati di dimensioni minori anche alle 100 bp: risultando pertanto particolarmente adatti ad analizzare DNA proveniente da tracce biologiche altamente degradate. Tuttavia, il limite principale nel loro utilizzo in ambito forense è connesso con il fatto che sono marcatori bi-allelici, diversamente dagli *STRs* che sono multiallelici, e pertanto la difficoltà di impiego nell'identificazione personale risiede nell'elevato numero di polimorfismi necessario esaminare al fine di ottenere un potere identificativo paragonabile a quello degli *STRs*. L'utilizzo di tale tipo di marcatore è stato recentemente incrementato a causa dell'introduzione delle nuove tecnologie *NGS* che permettono l'analisi, in un'unica reazione, di un numero molto elevato di *SNPs*.

5.3 Il DNA mitocondriale

Lo 0,25% del genoma umano è rappresentato dal DNA mitocondriale (mtDNA), molecola circolare a doppio filamento superavvolta, che, a differenza del DNA nucleare, non è associato ad istoni o ad altre proteine ed è privo di introni e sequenze ripetute. Ciascun mitocondrio può contenere da 1 a 15 copie di mtDNA, con una media di 4-5 copie per cellula, e siccome ogni cellula può contenere centinaia o addirittura migliaia di mitocondri a fronte della singola copia del DNA nucleare, il mtDNA è stato per anni il principale ed unico marcatore investigato negli studi sul DNA antico (*Cavelier et al. 2000*). Il DNA mitocondriale è costituito da circa 16.569 bp, anche se il numero di nucleotidi può variare nei diversi individui per la presenza di inserzioni o delezioni di una o più basi. La molecola possiede 37 geni, di cui 13 codificano proteine

implicate nel processo di fosforilazione ossidativa e nella produzione di energia, 2 geni codificano RNA ribosomiali (rRNA), 12S per la subunità piccola e 16S per la subunità grande, ed i restanti 22 geni codificano *RNA transfer* (tRNA) (Anderson et al. 1981). I geni che codificano proteine sono localizzati su entrambe le eliche e gli mRNA (RNA messaggeri), una volta sintetizzati, rimangono all'interno del mitocondrio e vengono tradotti sui ribosomi mitocondriali. Inoltre il mtDNA possiede una regione di controllo, non codificante, *D-loop* (*displacement loop*) lunga circa 1.122 bp, che mostra la maggiore variabilità fra gli individui in ragione del fatto che è una regione non codificante che svolge un ruolo importante nella regolazione e nell'attivazione dei processi replicativi e trascrizionali. Tale regione viene per lo più analizzata negli studi evolutivi e in ambito forense solo quando, in campioni altamente degradati, non è possibile riuscire ad ottenere un profilo genetico dall'analisi dei marcatori nucleari. La *D-loop* contiene al suo interno tre regioni che presentano il maggior grado di polimorfismo e che vengono analizzate nelle analisi forensi: la regione *HVRI* (*hypervariable region I*) che si estende dalla posizione 16.024 alla 16.365, la regione *HVRII* che si estende dalla posizione 73 alla 340 (*hypervariable region II*) e la regione *HVRIII* dalla posizione 438 alla 574 (*hypervariable region III*) (Figura 2).

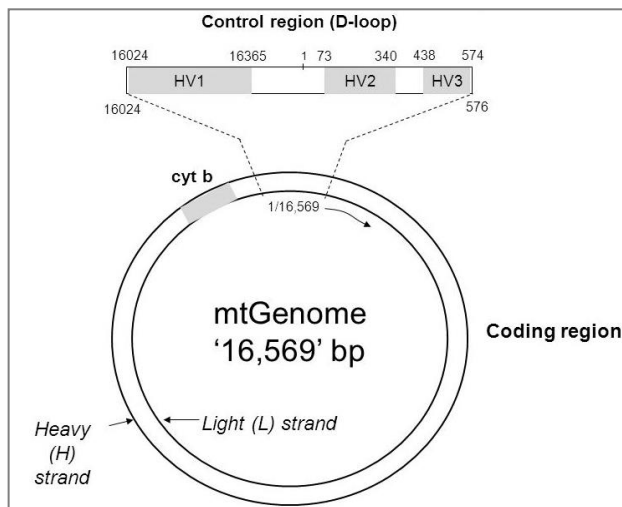


Figura 2: Il DNA mitocondriale e la *D-loop* suddivisa nelle sue tre regioni ipervariabili

Come già accennato, il DNA mitocondriale ha un'ereditarietà strettamente materna (*Giles et al. 1980*), le cellule uovo contengono circa 100.000 molecole di mtDNA mentre quelle spermatiche ne contengono solo dalle 100 alle 150; inoltre durante l'embriogenesi le cellule spermatiche perdono subito i loro mitocondri in quanto riconosciuti dai meccanismi di degradazione appena formati nell'embrione per la presenza di una molecola di ubiquitina (*Sutovsky et al. 2000*). Quindi indipendentemente dal fatto che un embrione sia maschio o femmina conterrà mitocondri derivanti esclusivamente dalla cellula uovo, quindi di derivazione materna.

6. L'analisi dei marcatori forensi

6.1 Tecnologia forense classica

Il flusso di lavoro di un laboratorio forense è riportato in Figura 3.



Figura 3: *Workflow* di un laboratorio forense

I campioni giunti in laboratorio, dopo un'accurata ispezione mirata all'individuazione delle tracce biologiche di interesse investigativo, vengono sottoposti ad estrazione del DNA in modo da permettere la separazione del DNA dalle altre componenti cellulari e dal substrato sul quale si trova adesa la traccia. Il DNA viene successivamente quantificato seguendo protocolli validati e specifici per campioni forensi. L'estratto viene quindi amplificato mediante *multiplex PCR* al fine di amplificare, in un'unica reazione, tutti i marcatori di interesse. Il prodotto di *PCR* viene poi fatto correre su sequenziatore capillare al fine di separare i prodotti di *PCR* in base al loro peso molecolare ed identificarli in base ai diversi fluorofori associati ai *primers*.

6.1.1 La *PCR multiplex*

La *PCR* o reazione a catena della polimerasi è una tecnica di biologia molecolare sviluppata negli anni '80 dal biochimico statunitense *Kary Mullis*. Tale tecnica permette di ottenere un gran numero di copie di un

dato frammento di DNA di interesse a partire da una quantità minima, anche da una singola molecola. Tale metodica si basa sulla proprietà naturale dell'enzima Taq DNA polimerasi, estratto da un batterio estremofilo, *Thermus aquaticus*, di essere termoresistente e di duplicare il DNA. L'utilità della PCR in campo forense è data dalla sua applicazione nell'analisi dei polimorfismi STRs.

Nel 1993 fu messo in commercio il primo kit di PCR multiplex prodotto dalla Promega. Tale kit permetteva l'amplificazione contemporanea di 3 marcatori (TH01, TPOX, CSF1PO) in un'unica reazione. La messa a punto della PCR multipla è risultata essere di notevole rilevanza per il biologo forense dal momento che riduceva il lavoro manuale e quindi i tempi di risposta e la quantità di DNA necessaria per ottenere un risultato utile. L'implementazione da parte delle aziende nella messa a punto di nuovi kits con un sempre maggior numero di marcatori analizzati in multiplex ha portato alla realizzazione di kits di ultima generazione per l'analisi STRs quali per es. il GlobalFiler® PCR Amplification Kit e l'AmpFISTR® NGM SElect™ PCR Amplification Kit (Thermo Fisher) ed il PowerPlex® Fusion Systems (Promega) che permettono la co-amplificazione di 24 loci in un'unica reazione. In tutti questi kits è compresa anche l'amplificazione del gene dell'Amelogenina, codificante per una proteina della matrice extracellulare espressa negli ameloblasti durante lo sviluppo dentale e coinvolta nel processo di produzione dello smalto. Questo gene è situato nella regione p22.2 sul braccio corto del cromosoma X, mentre sul cromosoma Y è situato nella regione p11.2 del braccio corto e presenta un'inserzione di 6 bp. In caso di individuo di

sesto femminile il *locus* Amelogenina é rappresentato sul tracciato elettroforetico da un solo picco relativo al cromosoma X (XX), in caso di individuo di sesso maschile sono presenti due picchi (uno relativo al cromosoma X e l'altro all'Y). (*Picozzi and Intini 2009*).

Tali *kits* utilizzano *primers* marcati con differenti fluorocromi (*dye*) per l'amplificazione in *multiplex* di marcatori di differenti dimensioni. Pertanto, per analizzare i prodotti di amplificazione, i frammenti vengono separati in funzione della loro lunghezza mediante elettroforesi capillare su sequenziatore automatico.

La tecnica di *PCR* viene impiegata anche per l'analisi classica forense degli *SNPs* poiché la prima fase dell'analisi è l'amplificazione della regione che circonda il *locus* in cui si trova lo *SNP* d'interesse, e poi, di conseguenza si procede con l'individuazione della mutazione puntiforme tramite una metodica di mini-sequenziamento detta *SNaPshot*. Quest'ultima tecnica tuttavia rende impossibile la sua applicazione al sequenziamento contemporaneo di un numero elevato di *SNPs*, pertanto per l'analisi di tali marcatori risulta più vantaggioso ricorrere all'utilizzo di tecniche di sequenziamento di ultima generazione che permettono l'analisi contemporanea di un numero di *SNPs* elevato.

6.1.2 L'elettroforesi capillare

I *kits* di *PCR* comunemente impiegati nella pratica forense consentono, come già detto, l'amplificazione simultanea di numerosi frammenti di DNA. Questi, trattandosi di *STRs*, sono costituiti da un numero di unità ripetute: alleli diversi presentano quindi differenti lunghezze degli

ampliconi generati. Di conseguenza per la loro analisi devono essere separati mediante un'opportuna tecnica che abbia una capacità di risoluzione tale da consentire di distinguere fra alleli che differiscono fra loro anche di una singola base ed in un *range* che vada dalle 100 alle 500bp. Il metodo utilizzato deve inoltre essere riproducibile per consentire il confronto dei risultati fra laboratori differenti. Per ottenere questa separazione fra le varie molecole presenti nella miscela di ampliconi prodotti dalla reazione di *PCR* si sfrutta la proprietà del DNA di possedere una carica negativa: in presenza di un campo elettrico gli ioni vengono attirati dal polo di carica opposta, quindi nel caso degli acidi nucleici, dal polo positivo. Questo processo prende il nome di elettroforesi e si riferisce alla migrazione di cariche elettriche in un mezzo di separazione alle cui estremità è applicata una differenza di potenziale. Se sottoposto ad un campo elettrico quindi, il DNA migrerà in direzione dell'anodo a carica positiva, allontanandosi dal catodo (elettrodo negativo) in funzione della differenza di potenziale applicata: più elevato è il voltaggio, maggiore sarà la forza del campo elettrico sulla molecola e più veloce sarà di conseguenza il suo movimento. Il DNA presenta però una carica negativa per ogni unità nucleotidica, con una distribuzione uniforme di carica per unità di massa; la forza del campo elettrico esercitata su molecole di dimensioni differenti sarebbe quindi la stessa, per questo motivo per il processo di elettroforesi si usano "setacci molecolari" costituiti da matrici porose al fine di separare le molecole in base alla loro lunghezza. Queste matrici sono costituite da gel o soluzioni polimeriche che consentono alle molecole più corte di muoversi più

rapidamente attraverso i pori, rallentando invece quelle di dimensioni maggiori. Poiché il movimento di cariche attraverso un campo elettrico genera calore che porta a modificare la viscosità della matrice polimerica, alterando così la mobilità elettroforetica della molecola, l'elettroforesi deve essere condotta in un sistema in grado di dissiparlo. Per questo l'apparecchiatura per elettroforesi è costituita essenzialmente da tre componenti principali: un alimentatore che genera una differenza di potenziale, un mezzo di separazione, i cui pori devono essere di dimensioni idonee alle molecole da "setacciare" ed un termostato che permetta il controllo e la regolazione della temperatura. La conduzione uniforme e regolare della corrente attraverso il sistema elettroforetico è garantita da tamponi di corsa (soluzioni saline a bassa forza ionica), grazie al movimento dei propri ioni che migrano insieme a quelli del campione. La mobilità elettroforetica ovvero la velocità di migrazione è direttamente proporzionale alla carica dello ione e al campo elettrico applicato ed inversamente proporzionale alle sue dimensioni e alla viscosità della matrice porosa usata come mezzo di dissipazione.

La tecnica dell'elettroforesi capillare (CE) fu introdotta nei primi anni '80 e dal successivo sviluppo della strumentazione ha guadagnato in breve popolarità nel campo della biologia molecolare e in quello forense. Questa strumentazione è completamente automatizzata e consente di esaminare più lunghezze d'onda simultaneamente e quindi un elevato numero di *loci* che si sovrappongono in lunghezza, con un minimo consumo di campione da sottoporre alla corsa, importante privilegio di questa tecnica, utile principalmente per campioni forensi non ripetibili.

L'innovazione della CE risiede nell'impiego di un sottile capillare in silice fusa, rivestito per permettere di maneggiarlo senza romperlo, riempito di un polimero viscoso che funge da setaccio molecolare: il diametro ridotto del capillare (diametro interno di 50-100 μm) permette infatti di poter impiegare differenze di potenziale 10-100 volte superiori a quelle impiegate per elettroforesi su gel (generalmente di circa 300 V/cm), diminuendo notevolmente i tempi di corsa. Un potenziale troppo elevato porterebbe a un eccessivo surriscaldamento del sistema; questo calore può essere facilmente dissipato grazie all'elevato rapporto tra superficie e volume, garantito dalla sottile conformazione del capillare, e alla sua lunghezza (25-75 cm). Le estremità del capillare sono immerse in due serbatoi contenenti un tampone di corsa e in cui si trovano due elettrodi, responsabili della generazione del campo elettrico. I campioni vengono iniettati nel capillare elettrocineticamente, attraverso l'esposizione ad alto voltaggio per pochi secondi, o aspirati mediante l'applicazione di un'elevata pressione. Per la separazione, alle estremità del capillare viene applicata una differenza di potenziale che fa migrare le molecole del campione verso l'elettrodo di carica opposta in funzione della loro carica e massa. In realtà, sulla mobilità degli ioni in elettroforesi capillare entra in gioco anche un fenomeno definito flusso elettro-osmotico (EOF). Infatti all'interno del capillare tutti i soluti (cationi, anioni e neutri) vengono spinti per effetto di questo flusso verso il catodo; ciò è dovuto alla ionizzazione della silice costituente il capillare: i gruppi silanologici acidi, che rivestono le pareti del capillare, assumono infatti carica negativa che attira i cationi del tampone,

attirando di conseguenza per osmosi le molecole di acqua che costituiscono il tampone. Si crea in questo modo un flusso che fa sì che tutte le specie, indipendentemente dalla loro carica, migrino in direzione del catodo, in quanto questo flusso risulta essere più grande di almeno un ordine di grandezza della mobilità ionica. La migrazione delle molecole all'interno del capillare avverrà quindi in funzione di questo flusso EOF e della mobilità ionica delle molecole in direzione dell'elettrodo di carica opposta. Si avrà quindi che i cationi migreranno più rapidamente, in quanto la loro mobilità sarà frutto dell'effetto sommato del flusso elettro-osmotico e della loro mobilità ionica in direzione del catodo; le molecole prive di carica migreranno invece in direzione dell'elettrodo positivo per il solo effetto del flusso EOF, mentre gli anioni saranno rallentati nella corsa verso il catodo dalla loro mobilità ionica in direzione dell'anodo. L'elettroosmosi è un fenomeno altamente dipendente dalle variabili ambientali: il flusso EOF aumenta all'aumentare di pH, campo elettrico e temperatura, mentre diminuisce all'aumentare della concentrazione del tampone. Le piattaforme per elettroforesi capillare utilizzano capillari rivestiti internamente che impediscono il flusso EOF grazie al mascheramento dei gruppi silanoliche carichi o polimeri che bloccano le cariche negative che si creano sulla superficie del capillare. Il flusso elettro-osmotico può infatti creare problemi nella riproducibilità delle separazioni di DNA variando la velocità delle molecole fra una corsa e l'altra. Grazie a questo tipo di capillari e all'impiego di questi polimeri, la separazione avviene solo per mobilità ionica, in funzione del solo rapporto massa/carica della molecola, garantendo la massima

riproducibilità a ogni sessione di corsa. Il segnale emesso dai fluorocromi, eccitati da un laser posto in prossimità dell'estremità anodica, viene registrato da un rivelatore attraverso una finestrella in corrispondenza del punto in cui manca il rivestimento sul capillare. Il rivelatore è costituito da una fotocamera CCD (*Charged-Coupled Device*), dispositivo ad accoppiamento di carica, ovvero un sensore in silicio in grado di rilevare la lunghezza d'onda della luce emessa dal fluorocromo eccitato. I fotoni che interagiscono col silicio danno origine a elettroni che vengono accumulati nelle celle di cui è costituito il dispositivo: maggiore sarà il numero di fotoni che colpisce la superficie della matrice di silicio, maggiore sarà l'accumulo di elettroni e di conseguenza l'altezza del segnale digitale in cui viene convertito. I dati vengono infine inviati a un computer che, mettendo in relazione il picco di fluorescenza con il tempo di migrazione, trasforma il segnale fluorescente in dato di lunghezza espresso in paia di basi (bp) o in sequenza nucleotidica. Questa tecnica consente di analizzare frammenti che si sovrappongono in dimensioni in quanto marcati con differenti fluorocromi che emettono fluorescenza a diverse lunghezze d'onda. In realtà, nonostante la differenza di emissione dei vari fluorocromi, resta comunque qualche sovrapposizione fra gli spettri di emissione. Per eliminare questo inconveniente, un algoritmo computerizzato, definito matrice, riconosce questa sovrapposizione e genera un unico picco riconducendolo all'emissione di un solo marcatore. La capacità di risoluzione e i tempi di corsa dipendono essenzialmente dal tipo di polimero impiegato, dalla sua concentrazione, dalle caratteristiche del

capillare e dal campo elettrico applicato: in generale la risoluzione di queste apparecchiature per CE deve essere di almeno 0.5bp per poter permettere di distinguere accuratamente ripetizioni parziali (micro-varianti alleliche) o alleli che differiscono fra loro di una sola base nucleotidica; polimeri più viscosi, così come capillari più lunghi, consentono una maggior risoluzione, a discapito però dei tempi di analisi, richiedendo tempi di corsa più lunghi.

6.1.3 Il sequenziamento SANGER

Il sequenziamento del DNA consiste nella determinazione della sequenza nucleotidica di una molecola di DNA. Una delle prime tecniche messe a punto ed ampiamente utilizzata in ambito forense in quella che viene chiamata “la prima generazione delle tecnologie di sequenziamento”, è il metodo *Sanger* (*Sanger and Coulson 1975*), detto anche metodo a terminazione di catena. Il protocollo prevede l’utilizzo dell’enzima DNA polimerasi e dei classici deossinucleotidi (dNTPs), con l’aggiunta di nucleotidi modificati, i dideossinucleotidi (ddNTPs), che si differenziano dai nucleotidi classici per l’assenza del gruppo ossidrilico sul carbonio 3’ necessario per formare il legame con il nucleotide successivo; l’aggiunta di un ddNTPs alla sequenza nascente impedisce il legame di qualsiasi altro nucleotide in quanto non si può formare il legame fosfodiesterico tra essi. Inizialmente il metodo *Sanger* prevedeva il clonaggio del campione di DNA da sequenziare all’interno di vettori e poi, una volta denaturata la doppia elica per ottenere delle molecole stampo a singolo filamento, queste venivano aggiunte alla miscela di reazione. Il DNA stampo a

singolo filamento veniva suddiviso in quattro provette e in ciascuna di esse venivano aggiunti: un *primer* specifico, complementare al filamento da sequenziare e marcato radioattivamente, la DNA polimerasi, che utilizzando il DNA come stampo era in grado di generare il filamento complementare, i dNTPs e uno dei quattro ddNTPs in quantità stechiometricamente inferiore. Il principio del metodo *Sanger* si basa sul fatto che, mentre la reazione procede, l'incorporazione casuale di un ddNTPs nel filamento nascente di DNA causa l'arresto dell'attività della polimerasi; in questo modo si possono generare una serie di frammenti della stessa sequenza, differenti per lunghezza, in corrispondenza del ddNTPs incorporato. I frammenti formati condividevano pertanto l'estremità iniziale ma differivano ciascuno rispetto agli altri nell'estremità 3' per un unico nucleotide. I frammenti marcati venivano caricati in 4 pozzetti diversi, fatti correre in un gel di poliacrilammide e le bande visualizzate mediante autoradiografia. La sequenza veniva ricostruita partendo dall'estremità 5' all'estremità 3' in base alla posizione delle bande nei vari pozzetti corrispondente al punto di terminazione della sintesi a livello di una delle quattro basi nucleotidiche per incorporazione di un ddNTPs. Successivamente è stata introdotta una variante al metodo originario in cui i frammenti venivano prodotti mediante amplificazione PCR evitando così di ricorrere al clonaggio in vettori. Inoltre un'ulteriore innovazione ha riguardato l'utilizzo di nucleotidi marcati ciascuno con un diverso fluorocromo con specifica lunghezza d'onda di emissione. Questo ha consentito di poter far correre tutti i frammenti in un'unica corsa e di automatizzare la lettura della

sequenza grazie all'utilizzo di una CCD camera atta a rilevare le differenti emissioni di fluorescenza ed a produrre l'elettroferogramma, un grafico costituito da una serie di picchi di quattro colori diversi (quanti sono i ddNTPs) che identificano ciascuno una determinata base. Nonostante il metodo *Sanger* sia stato per anni il principale strumento per la determinazione della sequenza di DNA di campioni antichi, forensi e di interi genomi (come nel caso di genomi batterici) e nonostante sia ancora utilizzato in alcuni ambiti, tuttavia la necessità di rispondere alle esigenze dettate dallo studio di genomi sempre più complessi ha portato alla necessità di sviluppare metodi automatizzati che consentano di analizzare un maggior numero di campioni in parallelo riducendo i tempi di analisi e ottenere al contempo maggiori informazioni. Lo sviluppo delle tecniche di sequenziamento di nuova generazione NGS ha permesso di incrementare enormemente la resa del sequenziamento.

6.2 La tecnologia NGS

Con il termine *NGS* si intende il sequenziamento in parallelo e massivo di molecole di DNA frammentate e amplificate in modo clonale. Queste piattaforme lavorando in maniera massiva e parallela consentono l'analisi di milioni o addirittura miliardi di paia di basi in poche ore, riducendo il tempo ed i costi di analisi rispetto alla tecnologia *Sanger*. Nonostante in commercio siano disponibili diverse tipologie di piattaforme *NGS*, tutte prevedono un *workflow* simile che consta di tre passaggi principali:

1. Preparazione della *library*: in questa fase alle estremità dei frammenti di DNA vengono legati degli adattatori cioè corti oligonucleotidi a sequenza nota su cui si potranno legare i *primers* nella successiva fase di amplificazione e sequenziamento;
2. Amplificazione della *library*: la procedura è differente a seconda del tipo di piattaforma utilizzata; in ogni caso, lo scopo di questa fase è quella di generare dei *clusters*, cioè un insieme di copie identiche di ciascun frammento di DNA in modo tale che il segnale generato sia sufficientemente intenso da poter essere rilevato dagli strumenti a disposizione;
3. Sequenziamento: il protocollo adottato differisce fra le varie piattaforme, così come l'*output*, cioè il numero di basi sequenziate per ogni corsa. Il sequenziamento produce una serie di *reads* che coprono in maniera più o meno completa l'intera sequenza di ciascun campione. Tali *reads* vengono analizzate mediante il ricorso a specifici *software* bioinformatici.

Questo tipo di sequenziamento può essere effettuato direttamente su tutto il DNA genomico estratto da un campione forense, nel caso del metodo *shotgun*, consentendo l'analisi di tutto il materiale genetico presente in un estratto, oppure su specifiche regioni di DNA selezionate tramite PCR o tramite il metodo *capture* con impiego di sonde oligonucleotidiche complementari specifiche alla regione d'interesse. Le sonde sono rappresentate da sequenze di RNA o DNA capaci di appaiarsi mediante ibridazione alla sequenza bersaglio, complementari a quello che si vuole

catturare/selezionare. Comunque venga effettuata, l'analisi di sequenziamento ultramassivo in parallelo permette di generare dei frammenti di basi dette *reads* che vengono riassemblati usando come stampo un DNA noto (*resequencing*), oppure riassemblati anche in mancanza del genoma di riferimento (*de novo sequencing*).

Ad oggi in commercio sono disponibili diverse piattaforme che permettono di ottenere, a seconda delle macchine utilizzate fino ad un *output* di 6Tb. Le aziende tuttavia hanno messo in commercio anche i cosiddetti “sequenziatori da banco”, apparecchiature di più modeste dimensioni che possono essere adottati anche dai laboratori più piccoli. Fra questi, si possono annoverare per esempio la *Ion S5* (*Life Technologies*) ed il *MiSeq FGx* (*Illumina*), che offrono un *throughput* (quantità di dati ottenuti) che va da 0.3 a 15 Gb per frammenti di 200bp nel caso dello *Ion S5* e da 540 Mb a 15 Gb nel caso del *MiSeq*.

6.2.1 La tecnologia NGS ILLUMINA

La tecnologia *MiSeq* *Illumina* si avvale della tecnologia SBS (*Sequencing by Synthesis*); la libreria a DNA con gli adattatori legati alle estremità del filamento viene denaturata per ottenere filamenti a singola elica che vengono immobilizzati in una *flowcell*, costituita da una *slide* otticamente trasparente con 8 canali (*lanes*) su cui sono legati gli oligonucleotidi di ancoraggio complementari. Si ha, a questo punto, la formazione del *cluster generation*, cioè un insieme di frammenti clonali di sequenza identica (possono essere generate più di 1000 copie del frammento d'interesse), mediante *bridge-PCR*, una PCR a ponte allestita su supporto

solido in cui il filamento di DNA si ripiega a ponte fra due oligonucleotidi di ancoraggio adiacenti. Nella fase di sequenziamento i *clusters* vengono denaturati e ibridati al 3' con il *sequencing primer*, la cui sequenza è complementare a quella dell'adattatore. Ai canali della *flowcell* viene aggiunta poi la DNA polimerasi e tutti e quattro i nucleotidi contemporaneamente. Si tratta di nucleotidi CTR (*cyclic reversible terminators*) che presentano cioè un blocco a livello del gruppo ossidrilico in 3' in maniera tale che possa essere incorporato un solo nucleotide alla volta. Il *laser* eccita il CTR e l'emissione di fluorescenza viene captata e registrata da una CCD camera. Il blocco in 3' del nucleotide può però essere rimosso chimicamente e un nuovo nucleotide può essere incorporato per procedere con l'estensione del filamento (Figura 4). I cicli di incorporazione ed emissione vengono stabiliti dall'operatore in funzione della lunghezza dei frammenti di DNA e si possono produrre 25 milioni di *reads* di lunghezza 2x300 bp ciascuna fino a 15 Gb di dati/run. Il vantaggio nell'utilizzo di questa tecnologia risiede nella possibilità di accorpare tutte le singole fasi del processo (dalla generazione dei cluster all'analisi dei dati) in un unico strumento e di completarle nell'arco di una giornata (sono necessarie circa 8 ore). Una piattaforma di questo tipo, ampiamente utilizzata per il sequenziamento di genomi microbici, ben si adatta allo studio del DNA antico grazie alla capacità di estrapolare dati di sequenza anche a partire da frammenti di DNA molto piccoli e del DNA forense per la sensibilità e per la capacità di analizzare diversi marcatori in un'unica corsa.

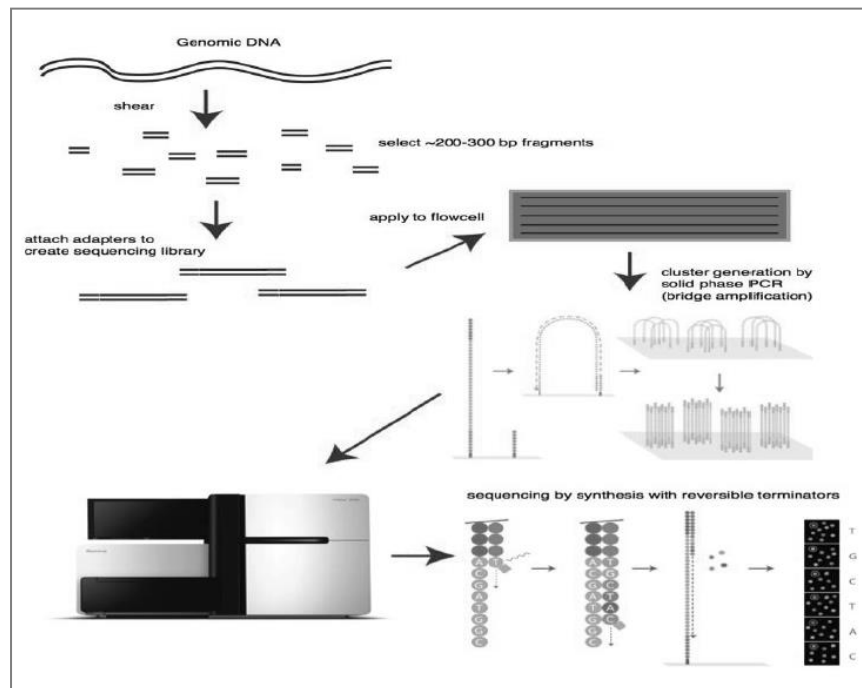


Figura 4: Fasi del processo di sequenziamento con piattaforma MiSeq Illumina.

6.2.2 La tecnologia *NGS ION TORRENT*

La tecnologia *Ion Torrent* è considerata la prima tecnica di sequenziamento che non prevede l'impiego della fluorescenza, bensì traduce un segnale chimico, dato da una differenza di potenziale, in segnale digitale. Come si può osservare in Figura 5, la strumentazione è composta da un chip a tre strati.

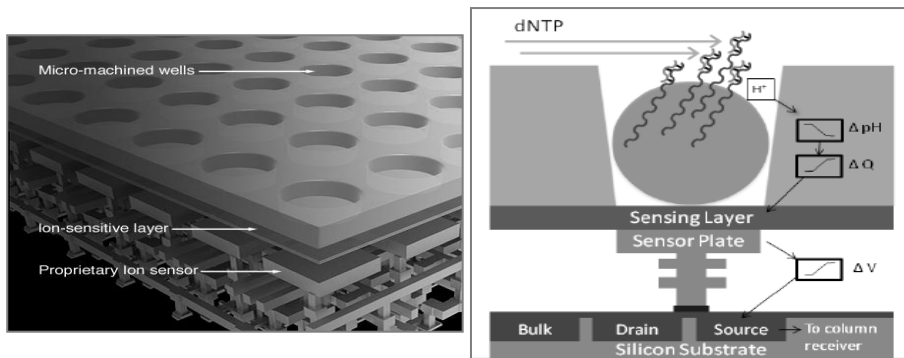


Figura 5: Struttura del semiconduttore Ion Torrent; Schema in sezione di un singolo pozzetto del chip semiconduttore contenente la Ion Sphere™ con associato il campione di DNA. Quando un nucleotide viene incorporato, vi è il rilascio di un protone e la determinazione di un cambio di pH. Il segnale chimico viene così tradotto in segnale digitale.

Il primo strato presenta numerosi micro-pozzetti contenenti ciascuno un filamento di DNA da sequenziare e l'enzima DNA polimerasi. Il secondo strato, sensibile agli ioni, e il terzo strato, che è un sensore ionico, costituiscono nel complesso il semiconduttore. Quando un dNTP è complementare al primo nucleotide del DNA stampo, all'interno del pozzetto avviene la sua incorporazione ad opera della DNA polimerasi con il conseguente rilascio di uno ione idrogeno. Il sensore ionico sottostante ai pozzetti rileva il cambiamento di pH della soluzione dato dal rilascio dello ione idrogeno e il sequenziatore traduce direttamente le informazioni codificate chimicamente (A, C, G, T) in informazioni digitali (0,1). Secondo il *workflow* di questa tecnologia (Figura 6), prima di procedere al sequenziamento il DNA stampo viene frammentato

mediante nebulizzazione o sonicazione così da poter procedere alla preparazione di una libreria di frammenti a doppio filamento lunghi diverse centinaia di basi presentanti alle estremità degli oligonucleotidi adattatori e degli indici.

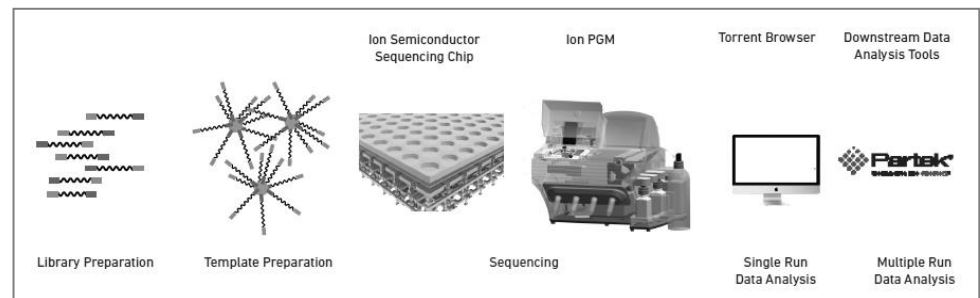


Figura 6: Schema delle fasi di lavoro con la tecnologia Ion Torrent

La successiva fase consiste nell'amplificazione clonale delle molecole di DNA, realizzata mediante la *PCR* in emulsione (*emPCR*), con impiego di nanosfere rivestite di oligonucleotidi complementari agli adattatori del frammento e grazie alla formazione di un'emulsione di acqua-olio. In Figura 7 è possibile seguire le diverse fasi della *PCR* emulsioclonale. Ciascuna gocciolina dell'emulsione cattura una singola nanosfera e una molecola di DNA, quindi all'interno di essa avviene la *PCR*. La molecola di DNA si denatura e uno dei due filamenti, detto *reverse strand*, si ibrida ad una delle due estremità della nanosfera per complementarità. All'altra estremità si ibrida un *primer* biotinilato, dopodiché ha inizio l'estensione del nuovo filamento, detto *forward strand*. Grazie ai cicli termici, si ottengono sulle singole superfici delle sfere

approssimativamente un milione di copie identiche tra loro di ogni frammento di DNA, generando localmente una genoteca di quel singolo frammento.

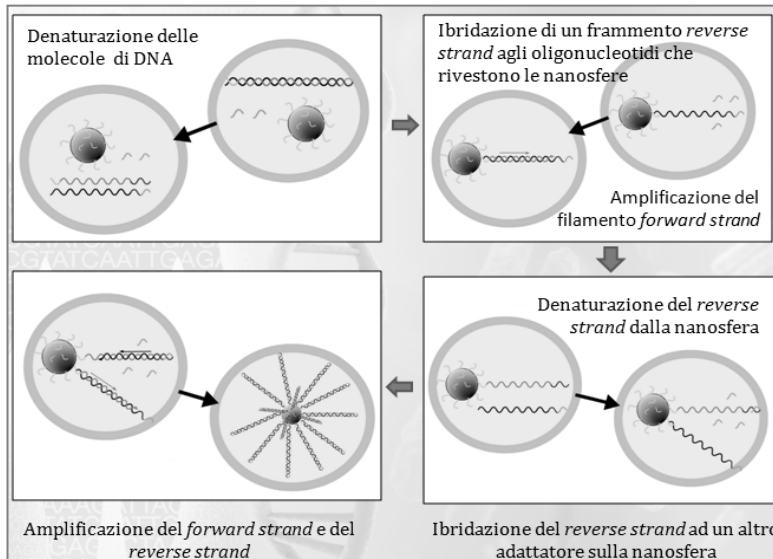


Figura 7: PCR emulsioclonale

Il punto critico di questa tecnica è riuscire a calcolare il giusto rapporto tra il numero delle sfere ed il numero di molecole del campione, in modo da evitare che più molecole diverse si leghino ad una stessa sfera, producendo quindi segnali non univocamente interpretabili in fase di sequenziamento; inoltre è da evitare che un numero troppo elevato di sfere rimangano non legate determinando, invece, una diminuzione della capacità di sequenziamento. Dopo l'amplificazione, l'emulsione viene rotta e si utilizzano delle biglie marcate con streptavidina per legare la

biotina con cui sono marcati i *primers* dei frammenti amplificati sulle nanosfere. In questo modo, grazie all'impiego di sostanze denaturanti le nanosfere con il DNA legato vengono separate da quelle a cui non sono legati frammenti. A questo punto il campione è pronto per essere sequenziato. I principali vantaggi di questa nuova tecnologia sono la velocità di sequenziamento e i bassi costi. Il limite di questa tipologia di analisi, risiede invece nella difficoltà di traduzione del segnale rilasciato in seguito ad una lunga e consecutiva ripetizione di uno stesso nucleotide, quando cioè nel frammento è presente una sequenza più o meno lunga dello stesso nucleotide.

6.2.3 Metodi di *target enrichment* e sequenziamento *shotgun*

Per lo studio del DNA degradato e presente in basso numero di copie sono disponibili differenti approcci di sequenziamento, la scelta di un particolare tipo di approccio o di un altro dipende dall'obiettivo del progetto e dalla quantità e qualità del materiale di partenza (*Rizzi et al. 2012*).

L'approccio utilizzato in questo lavoro di tesi è sia il *target enrichment capture* che si basa sull'utilizzo di sonde specificatamente designate per legarsi ad una precisa sequenza di DNA, garantendo contemporaneamente l'arricchimento del campione (*target enrichment*) ed il recupero di informazioni dal *pattern* di misincorporazione alle estremità 3' e 5' del filamento che l'arricchimento tramite *PCR*. Nel *target enrichment capture* le sequenze d'interesse vengono catturate mediante sonde complementari a specifiche regioni delle librerie,

evitando in tal modo di sequenziare interi genomi o sequenze non d'interesse, aumentando l'efficienza del sequenziamento e riducendo al tempo stesso i costi (Ávila-Arcos *et al.* 2011). I metodi di cattura disponibili sfruttano la capacità delle tecnologie di nuova generazione di sequenziare simultaneamente diversi campioni, adottando una strategia definita *multiplexing*. Affinché ciò sia possibile, alle librerie vengono aggiunti oligonucleotidi di sequenza nota alle estremità dei frammenti che prendono il nome di indici nella tecnologia Illumina e che identificano in modo univoco tutti i frammenti di uno stesso campione.

Il primo metodo di *target enrichment* utilizzato nel campo del DNA antico/degradato è stato la *Primer Extension Capture* (PEC) che utilizzava sonde biotilate che si legavano ad un determinato frammento di interesse, una DNA polimerasi che sintetizzava il filamento complementare e biglie ricoperte di streptavidina per il recupero delle sonde con attaccato il DNA sfruttando il legame streptavidina-biotina (Briggs *et al.*, 2009). Per ridurre i costi della PEC, Maricic e colleghi nel 2010 hanno dimostrato come sia possibile produrre “in casa” le sonde partendo dall'amplificazione di DNA moderno. Utilizzando due *PCR long-range* è possibile produrre due amplificati che sovrapposti possono ricoprire l'intero genoma mitocondriale. I due prodotti dell'amplificazione vengono in seguito frammentati mediante sonicazione e legati a molecole di biotina ad entrambe le estremità.

Oltre al sequenziamento mediante cattura sono disponibili altri due approcci di sequenziamento; il primo si basa sulla selezione di DNA *target* mediante *PCR* che, grazie all'utilizzo di *primers* specifici,

consente di selezionare determinate regioni di interesse del DNA e di produrre milioni di copie della stessa sequenza. Per ottimizzare tale procedura e produrre risultati soddisfacenti è cruciale il *designing* dei *primers* complementari alle regioni della sequenza di interesse che si vuole amplificare. Questo tipo di approccio può essere adottato quando la sequenza del frammento di interesse è nota. In particolare il metodo di *target enrichment* tramite *PCR* viene usato quando ad esempio l'obiettivo è identificare determinati polimorfismi presenti in una regione del DNA, come ad esempio gli *SNPs* utilizzati come *markers* nella determinazione degli aplotipi o per scopi identificativi. Oltre alla classica *PCR* è possibile effettuare una *multiplex PCR* utilizzando, nella stessa miscela di reazione, coppie di *primers* specifiche per diverse regioni del DNA. Una reazione di questo tipo genera differenti ampliconi del DNA target che possono essere sequenziati simultaneamente. Quando si vuole utilizzare un approccio di questo tipo devono essere considerati vantaggi e svantaggi; certamente il principale vantaggio, oltre all'elevata specificità, è rappresentato dalla generazione di un'elevata quantità di copie del frammento che rappresenta l'*input* ideale per la preparazione di *libraries* *NGS*. Tuttavia è necessario tener presente che le reazioni di *PCR* aumentano la percentuale di misincorporazioni all'estremità dei filamenti, che, come già detto, rappresentano un importante *marker* nella distinzione tra DNA antico e moderno. Un'altra modalità di sequenziamento è il metodo *shotgun* che viene eseguito direttamente su tutto il genoma, sia su materiale fossile che fresco. È stato il primo approccio ad essere impiegato nello studio del DNA antico nonostante, in

queste circostanze, la quantità di DNA endogeno sia notevolmente bassa rispetto all'eventuale contaminazione batterica e fungina. Questo rappresenta un forte limite per l'utilizzo di questa tecnologia che è per lo più indicata negli studi metagenomici, il cui obiettivo è l'identificazione di tutte le possibili specie all'interno di un campione.

Neither femur nor tooth: Petrous bone for identifying archaeological bone samples via forensic approach

Pilli E, Vai S, Caruso MG, D'Errico G, Berti A, Caramelli D.

Forensic Sci Int. 2017 Dec 24;283:144-149.

INTRODUZIONE

La capacità di rinvenire ed analizzare il DNA dai resti scheletrici rappresenta una delle sfide più significative per gli esperti di genetica forense e per gli antropologi molecolari. L'analisi del DNA dei resti umani è divenuta una pratica comune al fine di identificare i resti stessi e fornire informazioni genetiche utili sia per gli studi evolutivi che per le indagini sulle relazioni di parentela in contesti archeologici.

Le ossa insieme ai denti ed ai capelli sono infatti i campioni biologici più duraturi in natura e pertanto, talvolta, l'unico substrato biologico a disposizione degli esperti per le analisi genetiche. Come riportato in precedenza, le maggiori problematiche connesse con tale tipologia di campioni sono la scarsa quantità di molecole di DNA endogene, la degradazione del DNA e la possibile presenza nell'estratto di inibitori, cioè molecole che possono essere co-estratte con il DNA di interesse e che possono interferire con le successive fasi analitiche. Inoltre, quando si lavora con tale tipologia di campioni, un'ulteriore attenzione deve essere posta alle contaminazioni da parte di DNA moderno che possono verificarsi ad opera del personale addetto all'esumazione, per impropria conservazione o manipolazione del campione o ad opera degli studiosi di

antropologia e che possono rendere difficile l'ottenimento di profili *STRs* o sequenze autentiche.

Per tali motivi, vari sforzi sono stati compiuti negli anni dagli scienziati al fine di ottimizzare le metodiche analitiche (*Dabney et al. 2013; Rohland et al. 2007; Rohland et al. 2010; Jakubowska et al. 2012; Marshall et al. 2014; Coulson-Thomas et al. 2015; Zupanič Pajnič et al. 2016; Irwin et al. 2012; Ambers et al. 2016*) e prevenire le contaminazioni in campioni antichi (*Cooper et al. 2000*) ma meno lavoro invece è stato condotto sull'individuazione del migliore elemento scheletrico in termini di conservazione del DNA endogeno. *Hagelberg et al. (Hagelberg et al. 1991)* nel loro lavoro mettono in evidenza che non tutti gli elementi scheletrici possono contenere una quantità di DNA sufficiente per la tipizzazione e pertanto diversi ricercatori forensi hanno tentato di determinare quale elemento scheletrico sia superiore agli altri in termini di conservazione del DNA (*Barta et al. 2014; Edson et al. 2004; Leney 2016, 40:31–49; Milos et al. J 2007; Mundorff et al. 2009; Edson et al. 2009*). Nel 2009 *Edson et al.* hanno campionato differenti frammenti di cranio su cui è stata effettuata l'analisi del DNA mitocondriale. I risultati a cui sono giunti mettono in evidenza come la rocca petrosa sia un frammento osseo ottimale da cui recuperare DNA mitocondriale. La rocca petrosa è localizzata alla base del cranio fra l'osso sfenoide e le ossa occipitali e compone la parte endocranica dell'osso temporale in cui è ospitato l'organo dell'udito e dell'equilibrio. Come si deduce dal nome stesso, la rocca petrosa è la parte ossea più dura e più densa del corpo dei mammiferi (*Frisch et al. 1998*).

Recentemente *Kulstein et al.* (*Kulstein et al. 2017*) hanno dimostrato la possibilità di ottenere profili *STRs* da rocche petrose campionate in ambito forense. Grazie allo sviluppo della tecnologia *NGS* associata all'approccio *shotgun*, nel 2014, due differenti studi pubblicati su *Nature Communications* e *Nature* (*Gamba et al. 2014; Rasmussen et al. 2014*) hanno mostrato come la rocca petrosa possa fornire una buona resa in termini di DNA endogeno in campioni antichi. *Gamba et al.* hanno analizzato genomi completi e parziali dalle rocche petrose di 13 individui (tutti gli individui sono stati datati e vanno dal Neolitico all'età del bronzo), dai denti di quattro individui, dalle coste di due e dal metacarpale e metatarsale di altri due individui al fine di comparare la resa di DNA endogeno ottenuta dalla rocca petrosa agli altri distretti ossei provenienti dallo stesso individuo. I risultati ottenuti hanno messo in evidenza come la quantità di DNA ottenuta dalla rocca petrosa era superiore rispetto a quella ottenuta dai denti e dalle altre ossa a causa presumibilmente della sua alta densità strutturale associata ad una ridotta degradazione *post-mortem* e batterio mediata. Un anno più tardi, nel 2015, i risultati ottenuti da *Gamba et al.* sono stati confermati da *Pinhasi et al.* (*Pinhasi et al. 2015*) su campioni provenienti da contesti archeologici europei datati tra 10.000-1.800 cal. BP.

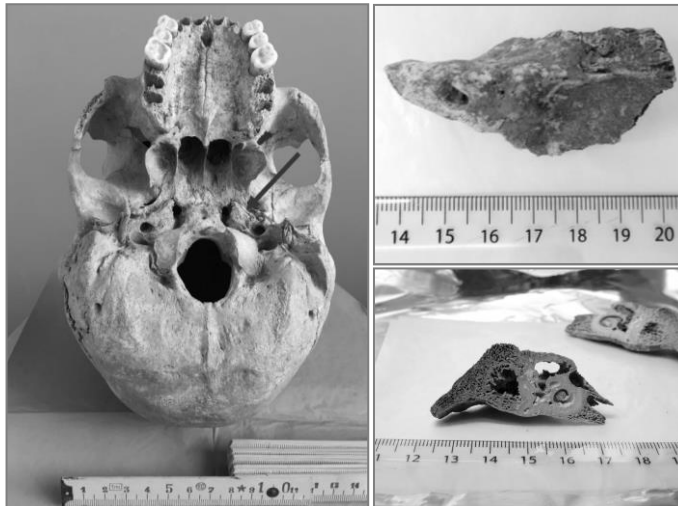
SCOPO DEL LAVORO

Sulla scia dei promettenti risultati ottenuti dall'analisi della rocca petrosa mediante l'utilizzo della tecnologia *NGS*, in questo lavoro, per la prima volta, abbiamo investigato la possibilità di utilizzare la rocca petrosa come elemento scheletrico utile per la tipizzazione *STR* in campioni ossei antichi sfruttando la metodologia di analisi forense classica PCR/CE (elettroforesi capillare). A tale scopo sono stati analizzati e confrontati tra loro tre differenti elementi scheletrici: femore (come suggerito dalla società internazionale di genetica forense, in qualità di “miglior” distretto osseo per le analisi genetiche -*Prinz et al. 2007*-), dente (in quanto in un lavoro di *Pilli et al. 2013* è risultato essere l'elemento scheletrico più refrattario alla contaminazione da parte di DNA esogeno rispetto alle altre ossa) e rocca petrosa provenienti da 13 individui del VI-VII secolo d.C. Il lavoro è consistito nella pulizia e polverizzazione dei distretti ossei oggetto di analisi e nell'estrazione del DNA in essi contenuto. Una volta quantificati, gli estratti sono stati sottoposti ad amplificazione e tipizzazione degli *STRs* al fine di comparare i risultati genetici ottenuti dai diversi elementi scheletrici investigati.

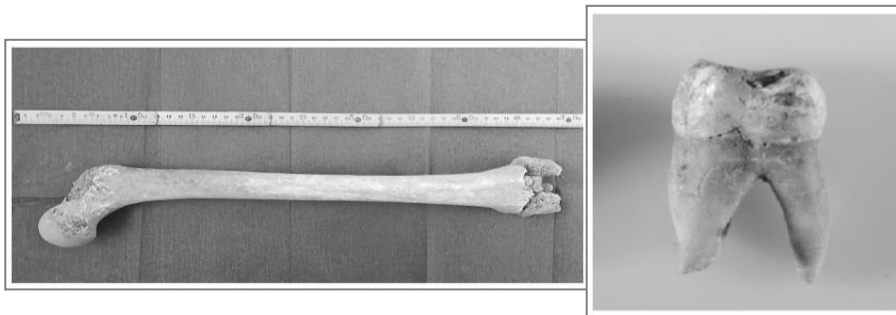
MATERIALI E METODI

1. I campioni

Trentanove elementi scheletrici sono stati campionati da tredici differenti individui. I campioni, selezionati da un sito archeologico Italiano, risalgono al VI-VII secolo d. C. Gli individui erano sepolti a terra in necropoli medievali strutturate. Tre differenti elementi scheletrici (femore, dente e rocca petrosa) per ciascun individuo (un esempio dei tre frammenti ossei è riportato in Figura 1) sono stati campionati per lo studio. Il campionamento è stato effettuato seguendo i protocolli utilizzati per i campioni antichi (*Fortea et al. 2008*). Alcune rocche petrose erano già scollegate dal cranio, mentre per i crani intatti, la rocca petrosa è stata tagliata verticalmente dalla parte inferiore del cranio con una lama e rimossa attraverso il forame magno. L'isolamento della parte interna dell'orecchio è stata condotta come in *Pinhasi et al. (Pinhasi et al. 2015)*.



A



B

Figura 1: Esempio di rocca petrosa (A), femore e dente (B) campionati per lo studio

2. L'ambiente di lavoro

Il lavoro sperimentale è stato condotto presso i Laboratori di Antropologia Molecolare dell'Università degli Studi di Firenze e presso la Sezione di Biologia del Reparto Investigazioni Scientifiche (RIS) di Roma. Il laboratorio di antropologia è dotato di due aree fisicamente

separate con accesso indipendente per eseguire le fasi di pre- e post-amplificazione al fine di prevenire il rischio di contaminazioni. Nella zona pre-amplificazione sono state eseguite le operazioni di pulizia, polverizzazione dei campioni e di estrazione del DNA. La zona dispone di una *dressing-room* in cui l'operatore deve indossare tuta, camice sterile, mascherina, cuffia per capelli, calzari e guanti sterili in modo da lavorare in condizioni di sterilità e ridurre al minimo il rischio di contaminazioni. Nell'area pre-amplificazione le superfici, gli strumenti, i materiali e la vetreria vengono lavate con ipoclorito di sodio e sottoposte a trattamento UV a 254 nm per almeno 45 minuti sia prima che dopo l'utilizzo. Particolare attenzione è rivolta anche ai reagenti che vengono suddivisi in aliquote sterili al fine di evitarne l'uso prolungato su un numero di campioni elevato. Controlli negativi sono stati aggiunti ai vari passaggi analitici.

3. Pulizia e polverizzazione dei campioni

Una volta giunti in laboratorio, i campioni sono stati accuratamente puliti rimuovendo 1-2 mm dello strato superficiale mediante l'utilizzo di un trapano odontoiatrico munito di punte metalliche precedentemente lavate con acido acetico al 30%, sciacquate in acqua bidistillata e sterilizzate sotto UV per almeno 45 minuti (*Adler et al 2011*). Una volta rimosso lo strato superficiale, i campioni sono stati sottoposti ad irraggiamento con UV per 45 minuti in *crosslinker (Biolink DNA crosslink)*.

Successivamente si è proceduto con la polverizzazione dei campioni: circa 50 mg di polvere d'osso per ciascun campione sono stati utilizzati per la successiva fase analitica.

4. Estrazione del DNA da polvere d'osso

La fase di estrazione del DNA si basa sulla progressiva solubilizzazione delle strutture cellulari che lo proteggono grazie al trattamento con soluzioni contenenti tensioattivi, agenti chelanti e enzimi quali le proteasi. Questi reagenti sono responsabili rispettivamente della distruzione delle membrane biologiche grazie alla solubilizzazione dei lipidi di membrana, dell'eliminazione dei sali di calcio che costituiscono la struttura ossea e della denaturazione delle proteine cellulari. L'estrazione è stata effettuata nella area PRE-amplificazione del laboratorio sotto cappa sterile. Al fine di monitorare l'effettiva condizione di sterilità del lavoro è stato incluso un controllo negativo (k-estrazione). Il protocollo di estrazione seguito è stato pubblicato da *Dabney* e collaboratori (*Dabney et al. 2013*) proposto come evoluzione del precedente protocollo pubblicato da *Rohland e Hofreiter* (*Rohland and Hofreiter 2007*). A differenza di quest'ultimo il protocollo proposto da *Dabney* necessita di una minore quantità di polvere d'osso di partenza (circa 50 mg contro i 150-250 mg richiesti nel protocollo precedente) che rappresenta un enorme vantaggio in quanto spesso i campioni a disposizione sono di dimensioni estremamente ridotte.

5. Quantificazione del DNA tramite Real Time PCR

La determinazione della quantità di DNA presente in un campione forense è fondamentale per la maggior parte delle successive analisi basate sulla *PCR* perché, se da un lato una quantità eccessiva di template potrebbe provocare la comparsa di picchi aggiuntivi o fuori scala per la tecnica di misurazione, dall'altro, una scarsa quantità di DNA potrebbe determinare la comparsa di fenomeni stocastici che potrebbero inficiare la reazione di *PCR* nonché la successiva interpretazione dei risultati della tipizzazione. Nella comune pratica di un laboratorio genetico-forense i risultati di quantificazione forniti dalla *Real Time PCR* (qPCR) assumono il ruolo di “linea di confine” tra la possibilità per un determinato campione di DNA di essere sottoposto o meno alle successive fasi analitiche, sulla base di una quantità di DNA compresa nell'intervallo ottimale indicato dal manuale d'uso dello specifico *kit* utilizzato. Per la quantificazione degli estratti di DNA è stata utilizzata la tecnica di PCR quantitativa avvalendosi del sistema *Applied Biosystems 7500 Real Time PCR*. Per l'analisi dei campioni di DNA è stato utilizzato il *kit Quantifiler® Trio DNA Quantification Kit (Thermo Fisher Scientific, Oyster Point, CA)*, che permette di rilevare simultaneamente la quantità totale di DNA umano amplificabile e la quota di DNA maschile eventualmente presente all'interno del campione. I risultati che si ottengono forniscono informazioni utili ad esempio per valutare se il campione contiene sufficiente quantità di DNA umano e/o DNA umano maschile per procedere con le successive fasi analitiche, stimare lo stato di degradazione del materiale genetico presente nel campione, valutare la

presenza di inibitori e determinare la quota di DNA maschile e femminile contenuta nell'estratto. La quantificazione con il *Quantifiler Trio* si basa sull'analisi di *loci target* a copia multipla per garantire un più elevato grado di sensibilità; in particolare, mediante l'utilizzo di *primers* e sonde specifici, si analizzano tre *target* caratteristici per il DNA umano: il *Large Autosomal* (LA) una regione di circa 200 bp che è usata come indicatore dello stato di degradazione del DNA, lo *Small Autosomal* (SA) una regione breve, compresa tra le 75 e le 80 bp, in grado di rilevare il DNA anche di campioni altamente degradati e l'*Y-Chromosome target*, una regione del cromosoma Y di circa 75-80 bp che permette di determinare la presenza di DNA genomico maschile all'interno del campione. Inoltre nel *kit* è compreso un ulteriore *target –IPC- (Internal PCR Control)* che viene aggiunto alla *mix* di reazione. IPC è una sequenza di DNA sintetica non presente in natura che, aggiunta a ciascun campione, funge da controllo interno positivo. Questo *target* risulta particolarmente utile per rilevare la presenza degli inibitori.

I dati di quantificazione ottenuti dal segnale dello *small autosomal target* permettono la determinazione della concentrazione dei campioni, mentre i dati del *large autosomal target* vengono utilizzati per valutare il livello di degradazione del DNA nei campioni. Poiché grandi frammenti di DNA sono più suscettibili alla degradazione di frammenti più piccoli, il rapporto fra i segnali provenienti dalle due sonde può essere utilizzato per determinare l'indice di degradazione (DI) secondo la formula matematica:

$$DI = CSA / CLA$$

dove CSA è la concentrazione dello *small autosomal target* nel campione e CLA è la concentrazione del *large autosomal target*. Il DI rappresenta pertanto un indicatore generale del livello di degradazione di un campione in cui maggiore è il valore dell'indice di degradazione maggiore è il livello di degradazione.

6. Amplificazione e Tipizzazione degli STRs

L'amplificazione del DNA è stata effettuata usando i due *kits* commerciali: *AmpFlSTR®NGM SElect™ PCR Amplification kit* (Applied Biosystems, Foster City, CA, USA) e *PowerPlex®ESX 17 System* (Promega, Madison, WI, USA). Questi *kits* di ultima generazione, comprendenti il set di marcatori *standard* utilizzati in Europa, amplificano rispettivamente:

- *AmpFlSTR NGM SElect* (Applied Biosystems): D10S1248, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D22S1045, D19S433, TH01, FGA, D2S441, D3S1358, D1S1656, D12S391, SE33 (Figura 2);

- *PowerPlex ESX 17* (Promega Corporation) amplifica gli stessi *loci* del *kit* precedente (Figura 2) ma le dimensioni dei marcatori sono invertite: i marcatori di grandi dimensioni nel primo *kit* sono di piccole dimensioni nel secondo e viceversa.

Table 1 AmpFSTR® NGM Select™ Kit loci and alleles			STR Locus	Label	Chromosomal Location ¹	Repeat Sequence ² 5' to 3'
D10S1248	10q26.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18	D18S51	Fluorescein	18q21.33 (59.1Mb)	AGAA (20)
vWA	12p13.31	11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24	D21S11	Fluorescein	21q21.1 (19.476Mb)	TCTA Complex (20)
D16S639	16q24.1	5, 8, 9, 10, 11, 12, 13, 14, 15	TH01	Fluorescein	11p15.5 (2.149Mb)	AATG (20)
D2S1338	2q35	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28	D3S1358	Fluorescein	3p21.31 (45.557Mb)	TCTA Complex
Amelogenin	X: p22.1-22.3 Y: p11.2	X, Y	Amelogenin ³	Fluorescein	Xp22.1-22.3 and Y	NA
D8S1179	8q24.13	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	D16S539	JOE	16q24.1 (84.944Mb)	GATA
D21S11	21q11.2-q21	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38	D2S1338	JOE	2q35 (218.705Mb)	TGCC/TTCC
D18S61	18q21.33	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	D1S1656	JOE	1q42 (228.972Mb)	TAGA Complex
D22S1045	22q12.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	D10S1248	JOE	10q26.3 (130.567Mb)	GGAA
D19S433	19q12	9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2	FGA	TMR-ET	4q28 (155.866Mb)	TTTC Complex (20)
TH01	11p15.5	4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3	D8S1179	TMR-ET	8q24.13 (125.976Mb)	TCTA Complex (20)
FGA	4q28	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2	vWA	TMR-ET	12p13.31 (5.963Mb)	TCTA Complex (20)
D2S441	2p14	9, 10, 11, 11.3, 12, 13, 14, 15, 16	D22S1045	TMR-ET	22q12.3 (35.779Mb)	ATT
D3S1358	3p21.31	12, 13, 14, 15, 16, 17, 18, 19	SE33	CXR-ET	6q14 (89.043Mb)	AAAG Complex
D1S1656	1q42.2	9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3	D19S433	CXR-ET	19q12 (35.109Mb)	AAGG Complex
D12S391	12p13.2	14, 15, 16, 17, 18, 19, 19.3, 20, 21, 22, 23, 24, 25, 26, 27	D12S391	CXR-ET	12p12 (12.341Mb)	AGAT/AGAC Complex
SE33	6	4.2, 6.3, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20.2, 21, 21.2, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35, 35.2, 36, 37	D2S441	CXR-ET	2p14 (68.214Mb)	TCTA

Figura 2: I 17 loci amplificati dal kit NGM Select a sinistra e PowerPlex ESX17 a destra. In tabella sono mostrate: nel kit NGM le posizioni sul cromosoma ed i possibili alleli ad ogni locus, mentre nel PowerPlex ESX il fluoroforo che marca il primer, la posizione dei loci sui cromosomi e il “core” dell’STR.

7. Il sequenziamento

I frammenti amplificati hanno un peso molecolare diverso per il variabile numero di ripetizioni contenute nella regione polimorfa e sono facilmente tipizzabili tramite elettroforesi capillare mediante marcatura fluorescente. L’elettroforesi capillare basata su rivelazione tramite fluorescenza, utilizzando la tecnologia della *multicolor detection* (a 5 fluorocromi), permette di analizzare loci multipli compresi quelli che hanno alleli che si sovrappongono in lunghezza, in una singola iniezione capillare. Dopo la corsa, i dati vengono elaborati da un computer e il risultato grafico dell’elettroforesi, detto elettroferogramma, viene restituito come picchi di

colore corrispondente a quello del fluorocromo usato per marcare il frammento di interesse.

Per l'analisi del DNA è stato utilizzato il sequenziatore automatico ABI PRISM 3500 *Genetic Analyzer* (Applied Biosystems) usando il polimero POP4, il capillare di 36 cm e le impostazioni predefinite dello strumento come consigliato dal produttore. I profili genetici sono stati determinati utilizzando il *software data collection 2 e GeneMapper® ID-X*. La soglia analitica è stata fissata a 50 RFU secondo validazione interna. Gli alleli sono stati considerati autentici quando presenti in entrambe le amplificazioni o se venivano chiaramente assegnati.

RISULTATI E DISCUSSIONE

Il DNA è stato estratto da tutti gli elementi scheletrici oggetto di studio e la concentrazione e degradazione del DNA è stata valutata. Nessun risultato è stato ottenuto sia per lo *small autosomal* che per il *large autosomal target* da tutti i campioni di femore e di dente analizzati ad eccezione di tre denti denominati 3, 15 e 45 in cui la concentrazione dello *small autosomal target* e dell'Y è risultata essere rispettivamente 10, 10 e 20 pg/μl (Figura 3).

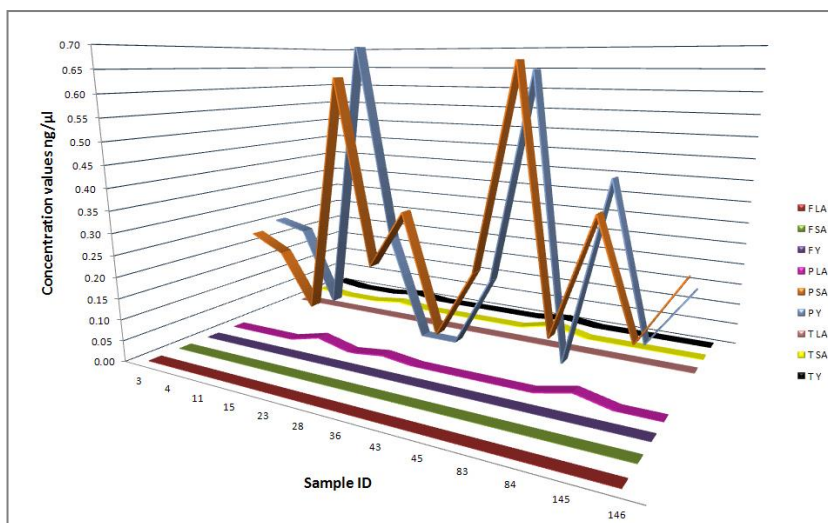


Figure 3: Risultati di quantificazione di tutti i campioni analizzati con il kit di quantificazione *Quantifiler Trio*. In ascisse il nome dei campioni, in ordinate la concentrazione espressa in ng/μl. F LA= femore *large autosomal target*; F SA= femore *small autosomal target* e F Y= femore cromosoma Y target; P LA= rocca petrosa *large autosomal target*; P SA= rocca petrosa *small autosomal target* e R Y= rocca petrosa cromosoma Y target; T LA= dente *large autosomal target*; T SA= dente *small autosomal target* e T Y= dente cromosoma Y target.

Come è possibile osservare dalla Figura 3, nessun risultato di quantificazione è stato ottenuto per la maggior parte dei campioni di femore e dente a causa presumibilmente della scarsa quantità del materiale genetico presente nei campioni associata all'elevata degradazione del materiale genetico. Presumibilmente per gli stessi motivi, anche la quantificazione del *large autosomal target* della maggior parte delle rocche petrose di quasi tutti i campioni, ad eccezione del 15, 28 ed 84 (20, 10 e 20 pg/ul rispettivamente) è risultata pari a 0. Al contrario, alti valori di quantificazione (valore medio di circa 261 pg/ul) sono stati ottenuti per lo *small autosomal target* di tutte le rocche petrose, in cui, in dieci campioni su tredici il valore di quantificazione è risultato essere maggiore di 100 pg/ul. Le differenze che sono state osservate nei valori di quantificazione ottenuti dalla sonda *small* nelle varie rocche petrose analizzate potrebbero essere dovute al campionamento di una parte della rocca petrosa differente a seconda dei campioni analizzati, come suggerito da *Pinhasi et al.* (*Pinhasi et al.* 2015). Simili risultati sono stati ottenuti dalla quantificazione del marcatore Y in cui i valori ottenuti variano in un *range* di 50-710 pg/ul per dieci dei tredici campioni quantificati. Tali risultati ci hanno portato ad ipotizzare di essere in presenza di dieci individui di sesso maschile. Nessun valore di quantificazione viene riportato (Figura 3) per le rocche petrose degli scheletri 28 e 36, i quali, in accordo con i precedenti studi antropologici, sono risultati appartenere presumibilmente ad individui di sesso femminile. Anche per quanto riguarda il campione 83, l'assenza di quantificazione del marcatore Y ci ha fatto ipotizzare di essere in

presenza di un individuo di sesso femminile anche se, essendo un'infante, l'analisi antropologica aveva dato risultati inconclusivi. Dal momento che nessun risultato è stato ottenuto dal *large autosomal target* per la maggior parte dei campioni analizzati, l'indice di degradazione è stato valutato solo per i campioni 15, 28 e 84. I valori stimati, come riportato nel manuale d'uso (*Quantifiler® HP and Trio DNA Quantification Kits User Guide*), pari a 31.5, 33 e 18.5 hanno evidenziato la presenza di una significativa degradazione del materiale genetico estratto. In seguito alla quantificazione, gli estratti sono stati amplificati e, come atteso, nessun risultato è stato ottenuto dal DNA estratto dai femori e dai denti ad eccezione dei denti 4, 11, 45 e 145 in cui solo 5, 3, 6 e 4 *loci*, non privi di discordanze (Tabella 1) se confrontati con il profilo ottenuto dalle rocche petrose, sono stati rispettivamente amplificati.

Campioni	R	F	D
3	14/17	0/17	0/17
4	17/17 ⁺	0/17	5/17* ⁺
11	13/17	0/17	3/17**
15	17/17	0/17	0/17
23	12/17	0/17	0/17
28	15/17	0/17	0/17
36	15/17	0/17	0/17
43	14/17	0/17	0/17
45	17/17	0/17	6/17 [§]
83	17/17	0/17	0/17
84	16/17	0/17	0/17
145	11/17	0/17	4/17 ^{§§}
146	16/17	0/17	0/17

Tabella 1: Profili *STRs* ottenuti dalle rocche petrose (R), dai femori (F) e dai denti (D) di ciascuno degli individui tipizzati. I risultati sono espressi come numero di *loci* positivamente tipizzati sul numero di *loci* totali analizzati. *discordanze osservate al *locus* D10S12468, amelogenina, D22S1045, D19S433 e D2S441; **discordanze osservate al *locus* D22S1045 e D3S1358; [§]discordanze osservate al *locus* D8S1179; ^{§§}discordanze osservate al *locus* D22S1045, D2S441 e D3S1358. ⁺Figure 4.

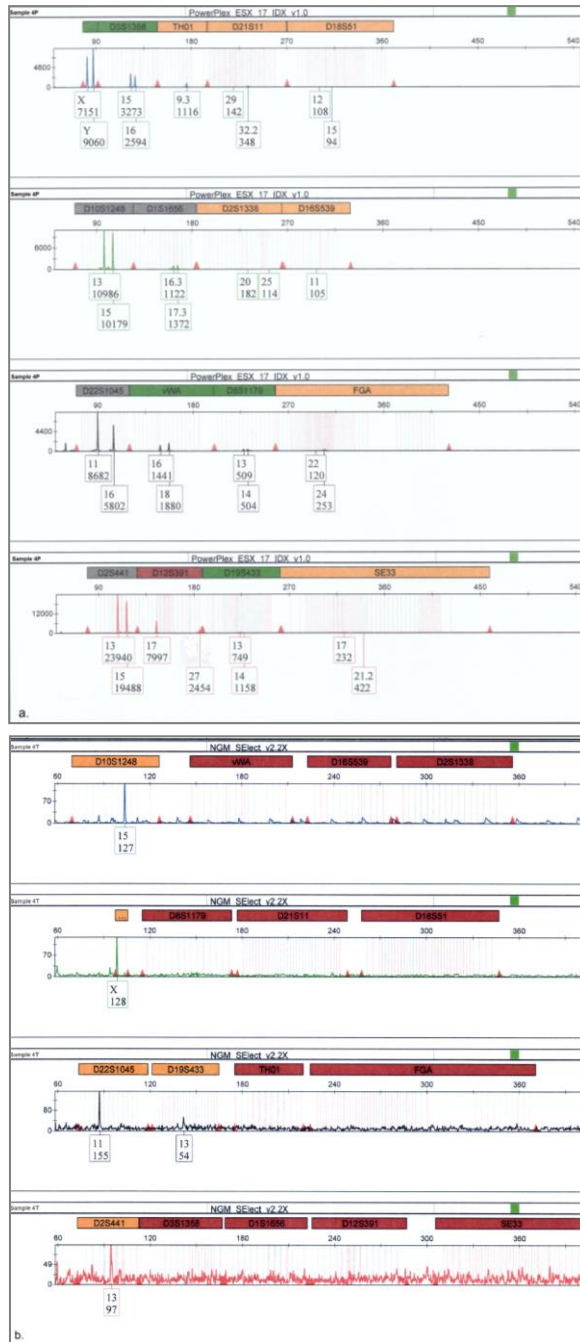


Figura 4: Profilo STR del campione 4: a. rocca petrosa e b. dente

Come è possibile osservare in Tabella 1, profili *STRs* completi, con un minimo di 13 *loci* tipizzati, sono stati osservati in 11 rocche petrose su 13. La percentuale media dei *loci* tipizzati è risultata pari a 88%, 0% e 8% rispettivamente per le rocche petrose, i femori ed i denti. Come atteso, quando si analizzano campioni altamente degradati, i profili ottenuti presentavano una curva di decadimento in cui è stato possibile osservare un decremento dell'altezza dei picchi dai *loci* più leggeri a quelli più pesanti. L'altezza dei picchi risulta infatti inversamente proporzionale alla lunghezza degli ampliconi (Lygo et al. 1994; Cotton et al. 2000; Wallin et al. 1998; Chung et al. 2004) (Figura 5).

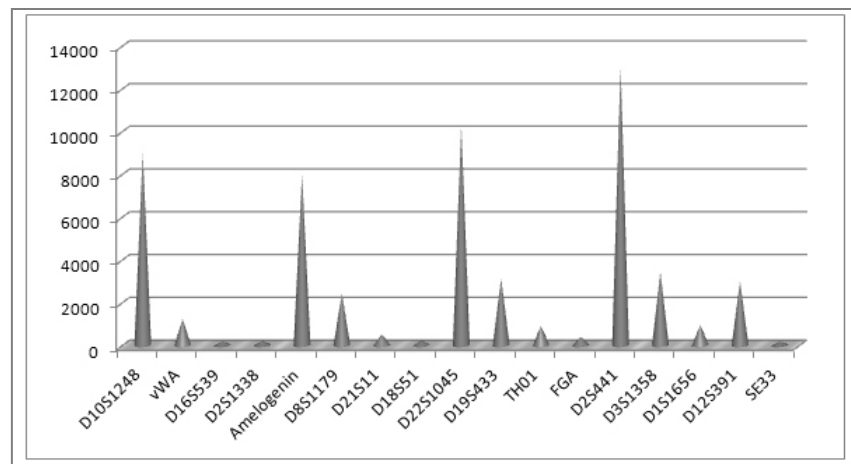


Figura 5: I valori di RFU media (unità di fluorescenza relativa) dei *loci* analizzati sono presentati in tabella. I *loci* sono stati organizzati in ordine discendente in accordo con i marcatori del kit *AmpFlSTR® NGM SElect™*.

Come è possibile osservare in Figura 5, i prodotti di PCR di circa 100 bp (D10S1248, D22S1045, e D2S441) hanno mostrato un segnale maggiore (picchi più alti) rispetto a quelli di circa 300-400 bp (D2S1338, D18S51 e SE33) a causa del fatto che in campioni degradati è più facile trovare frammenti di piccole dimensioni intatti rispetto a quelli di dimensioni maggiori. Il valore di RFU medio osservato nel *locus* D12S391 (amplicone di dimensioni maggiori), maggiore rispetto al *locus* D1S1656 di più piccole dimensioni, potrebbe essere spiegato con la diversa dimensione del marcatore D12S391 nei due *kits* utilizzati. Inoltre la media degli RFU di ciascun profilo è stata calcolata ed è stato osservato che variava in un *range* di 641-9.294 RFU per i profili ottenuti dalla rocca petrosa mentre in un *range* di 0-232,3 RFU per i profili dei denti. In accordo con i risultati di quantificazione, questi dati hanno messo in evidenza una maggiore quantità di DNA negli estratti delle rocche petrose rispetto a quella dei denti e dei femori. I risultati ottenuti in questo lavoro appaiono tuttavia discordanti rispetto a quelli ottenuti da *Kulstein et al.* (*Kulstein et al. 2018*) i quali, nel loro lavoro, erano riusciti ad ottenere profili *STRs* completi sia dalle rocche petrose che dalle ossa lunghe e dai denti campionati durante le attività investigative nel corso degli anni. La mancata tipizzazione dei femori e dei denti analizzati in questo lavoro potrebbe essere spiegata con l'età dei campioni come suggerito da *Sawyer et al.* (*Sawyer et al 2012*) per il DNA mitocondriale (...*However, although the average size of fragments retrieved from older specimens can be larger than those retrieved from younger specimens, over-all the total amounts of mtDNA retrievable from specimens*

decrease over time). In accordo con i nostri dati, *Rothe et al.* (*Rothe et al. 2015*) sono riusciti ad ottenere profili *STRs* completi da rocche petrose di tre dei quattro individui medievali campionati e solo tre profili parziali da femori e denti. Un discorso a parte deve essere fatto per le rocche petrose, le quali non solo non sembrano risentire dello scorrere del tempo (profili *STRs* sono stati ottenuti da tutte le rocche antiche analizzate) ma i valori di concentrazione media dei campioni antichi risultano maggiori rispetto a quelli dei campioni ossei forensi. Tali risultati, sorprendenti considerata l'età dei campioni raccolti (VI-VII secolo d.C.), potrebbero essere dovuti al fatto che la rocca petrosa si trova in una posizione protetta del cranio ed è una struttura scheletrica densa con una ridotta degradazione *post-mortem* e batterio mediata. Va inoltre tenuto in considerazione che *Rothe et al.* (*Rothe et al. 2015*), nel loro lavoro, non fanno alcun riferimento (“*bone powder was obtained intracranially with a rasp after the outer layer was rasped and discarded, because of potential contamination with modern DNA*”) all'utilizzo di una lama rotante per la separazione/apertura della rocca petrosa e l'isolamento della parte interna dell'osso (parte C) che *Pihnasi et al.* (*Pinhasi et al. 2015*) avevano dimostrato essere la parte “migliore” in termini di concentrazione di DNA endogeno. Riassumendo il nostro studio ha dimostrato la fattibilità della tipizzazione *STR* di campioni ossei altamente degradati/antichi (rocche petrose). Inoltre l'utilizzo di due differenti *kits* non solo ci ha permesso di confermare i risultati ottenuti ma anche, in alcuni casi, di aumentare il numero di *loci* tipizzati. Infine la quantificazione e l'analisi *STR* hanno reso possibile l'identificazione del

sexo degli individui campionati, confermando quanto supposto dalle indagini antropologiche.

CONCLUSIONI

Le ossa rappresentano il materiale biologico che dura più a lungo nel tempo e che risulta utile per l'identificazione umana quando le altre tracce biologiche non sono più disponibili. Tuttavia, a causa della degradazione del DNA in esse contenuto associata alla scarsa quantità di materiale genetico a disposizione, l'identificazione di resti umani (individui provenienti da sepolture antiche rinvenuti in contesti archeologici) mediante la tipizzazione degli *STRs* è fortemente dipendente dalla capacità di isolare quantità sufficienti di DNA dai campioni. Dal confronto di tre diversi elementi scheletrici prelevati dallo stesso scheletro antico e quindi sottoposti allo stesso *stress* ambientale, per la prima volta su campioni antichi datati VI-VII secolo d.C., è stato possibile dimostrare che l'osso petroso è l'elemento scheletrico migliore ai fini identificativi in termini di quantità e qualità di DNA isolato rispetto al femore e al dente presumibilmente a causa dell'elevata densità di questo elemento scheletrico associata alla resistenza al danneggiamento del DNA e alla sua "protezione" dalla degradazione. La presenza di una buona quantità di materiale genetico, nonostante l'elevata degradazione dello stesso, ci ha permesso di quantificare e successivamente identificare, mediante metodica forense classica, i profili *STRs* di ossa petrose antiche ottenendo profili *STRs* completi per quattro su tredici campioni ed un numero di *loci* tipizzati superiore a 11 per tutti i campioni. Particolare attenzione dovrebbe essere posta quando

l'osso petroso viene campionato dalla sua posizione endocraniale dal momento che il prelievo potrebbe portare alla rottura del cranio. Tuttavia, come effettuato in questo lavoro, è possibile rimuovere l'osso petroso senza modificare l'integrità del cranio tagliando la connessione tra osso petroso e temporale utilizzando una lama chirurgica e recuperando l'osso petroso dal forame magno lasciando solo un piccolo foro visibile dalla vista basale.

Da quanto emerge dal nostro studio, le ossa petrose antiche possono essere considerate una buona fonte di DNA e il DNA estratto da esse può essere utilizzato per scopi identificativi, sfruttandole per esempio come substrato biologico nei *cold case* in cui il DNA è compromesso sia in termini qualitativi che quantitativi così come negli studi archeologici. Va inoltre tenuto in considerazione il fatto che la tipizzazione *STRs* delle rocche petrose potrebbe rappresentare un modo economico e veloce per investigare le relazioni di parentela in contesti archeologici e valutare il sesso degli individui quando il materiale scheletrico non è utile per stime morfometriche come nel caso dei neonati.

BIBLIOGRAFIA

- Adler, C.J. Haak, W. Donlon, D. Cooper, A.** 2011. Survival and recovery of DNA from ancient teeth and bones. *J Arch Sci* 38:956–964.
- Ambers, A. Turnbough, M. Benjamin, R. Gill-King, H. King, J. Sajantila, A. Budowle, B.** 2016. Modified DOP-PCR for improved STR typing of degraded DNA from human skeletal remains and bloodstains. *Leg Med (Tokyo)* 18:7–12.
- Barta, J.L. Monroe, C. Crockford, S.J. Kemp, B.M.** 2014. Mitochondrial DNA preservation across 3000-year-old northern fur seal ribs is not related to bone density: Implications for forensic investigations. *Forensic Sci Int* 239:11–8.
- Chung, D.T. Drábek, J. Opel, K.L. Butler, J.M. McCord, B.R.** 2004. A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets. *J Forensic Sci* 2004, 49:733–40)
- Cooper, A. Poinar, H.N.** 2000. Ancient DNA: do it right or not at all. *Science* 289:1139
- Coulson-Thomas, Y.M. Norton, A.L. Coulson-Thomas, V.J. Florencio-Silva, R. Ali, N. Elmrghni, S. Gil, C.D. Sasso, G.R.S. Dixon, R.A. Nader, H.B.** 2015. DNA and bone structure preservation in medieval human skeletons. *Forensic Sci Int* 2015, 251:186–94.
- Cotton, E.A. Allsop, R.F. Guest, J.L. Frazier, R.R. Koumi, P. Callow, I.P. Seager, A. Sparkes, R.L.** 2000. Validation of the ampF/STR

SGM plus system for use in forensic casework. *Forensic Sci Int* 112:151–61.

Dabney, J. Knapp, M. Glocke, I. Gansauge, M-T. Weihmann, A. Nickel, B. Valdiosera, C. García, N. Pääbo, S. Arsuaga, J-L. Meyer, M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* 110:15758–63.

Edson, S.M. Ross, J.P. Coble, M.D. Parsons, T.J. Barritt, S.M. 2004. Naming the Dead - Confronting the Realities of Rapid Identification of Degraded Skeletal Remains. *Forensic Sci Rev* 16:63–90.

Edson, S.M. Christensen, A.F. Barritt, S.M. Meehan, A. Leney, M.D. Finelli, L.N. 2009. Sampling of the cranium for mitochondrial DNA analysis of human skeletal remains. *Forensic Sci Int Genet Suppl Ser* 2:269–270.

Fortea, J. de la Rasilla, M. García-Taberner, A. Gigli, E. Rosas, A. Lalueza-Fox, C. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidrón Cave (Asturias, Spain). *J Hum Evol* 55:353–7.

Frisch, T. Sørensen, M.S. Overgaard, S. Lind, M. Bretlau, P. 1998. Volume-referent bone turnover estimated from the interlabel area fraction after sequential labeling. *Bone* 22:677–82.

Gamba, C. Jones, E.R. Teasdale, M.D. McLaughlin, R.L. Gonzalez-Fortes, G. Mattiangeli, V., Domboróczki, L. Kóvári, I. Pap, I. Anders, A. Whittle, A. Dani, J. Raczky, P. Higham, T.F.G.

- Hofreiter, M. Bradley, D.G. Pinhasi, R.** 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5:5257.
- Hagelberg, E. Gray, I.C. Jeffreys, A.J.** 1991. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* 352:427–9.
- Irwin, J.A. Just, R.S. Loreille, O.M. Parsons, T.J.** 2012. Characterization of a modified amplification approach for improved STR recovery from severely degraded skeletal elements. *Forensic Sci Int Genet* 6:578–87.
- Jakubowska, J. Maciejewska, A. Pawłowski, R.** 2012. Comparison of three methods of DNA extraction from human bones with different degrees of degradation. *Int J Legal Med* 126:173–8.
- Kulstein, G. Hadrys, T. Wiegand, P.** 2018. As solid as a rock—comparison of CE- and MPS-based analyses of the petrosal bone as a source of DNA for forensic identification of challenging cranial bones. *Int J Legal Med* 132(1):13-24.
- Leney, M.D.** 2016. Sampling Skeletal Remains for Ancient DNA (aDNA): A Measure of Success. *Hist Archaeol* 40:31–49.
- Lygo, J.E. Johnson, P.E. Holdaway, D.J. Woodroffe, S. Whitaker, J.P. Clayton, T.M. Kimpton, C.P. Gill, P.** 1994. The validation of short tandem repeat (STR) loci for use in forensic casework. *Int J Legal Med* 107:77–89.
- Marshall, P.L. Stoljarova, M. Schmedes, S.E. King, J.L. Budowle, B.** 2014. A high volume extraction and purification method for

recovering DNA from human bone. *Forensic Sci Int Genet* 2014, 12:155–60.

Milos, A. Selmanović, A. Smajlović, L. Huel, R.L.M. Katzmarzyk, C. Rizvić, A. Parsons, T.J. 2007. Success rates of nuclear short tandem repeat typing from different skeletal elements. *Croat Med J* 48:486–93.

Mundorff, A.Z. Bartelink, E.J. Mar-Cash, E. 2009. DNA preservation in skeletal elements from the World Trade Center disaster: recommendations for mass fatality management. *J Forensic Sci* 54:739–45.

Quantifiler® HP and Trio DNA Quantification Kits User Guide.

Rohland, N. Hofreiter, M. 2007. Comparison and optimization of ancient DNA extraction. *Biotechniques* 2007, 42:343–52.

Rohland, N. Siedel, H. Hofreiter, M. 2010. A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour* 10:677–83.

Pilli, E. Modi, A. Serpico, C. Achilli, A. Lancioni, H. Lippi, B. Bertoldi, F. Gelichi, S. Lari, M. Caramelli, D. 2013. Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. *PLoS One* 8:e52524.

Pinhasi, R. Fernandes, D. Sirak, K. Novak, M. Connell, S. Alpaslan-Roodenberg, S. Gerritsen, F. Moiseyev, V. Gromov, A. Raczky, P. Anders, A. Pietrusewsky, M. Rollefson, G. Jovanovic, M. Trinhhoang, H. Bar-Oz, G. Oxenham, M. Matsumura, H.

- Hofreiter, M.** 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* 10(6):e0129102.
- Prinz, M. Carracedo, A. Mayr, W.R. Morling, N. Parsons, T.J. Sajantila, A. Scheithauer, R. Schmitter, H. Schneider, P.M.** 2007. DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Sci Int Genet* 1:3–12.
- Rasmussen, M. Anzick, S.L. Waters, M.R. Skoglund, P. DeGiorgio, M. Stafford, T.W. Rasmussen, S. Moltke, I. Albrechtsen, A. Doyle, S.M. Poznik, G.D. Gudmundsdottir, V. Yadav, R. Malaspinas, A-S. White, S.S. Allentoft, M.E. Cornejo, O.E. Tambets, K. Eriksson, A. Heintzman, P.D. Karmin, M. Korneliussen, T.S. Meltzer, D.J. Pierre, T.L. Stenderup, J. Saag, L. Warmuth, V.M. Lopes, M.C. Malhi, R.S. Brunak, S. et al.** 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–9.
- Rohland, N. Hofreiter, M.** 2007. Ancient DNA extraction from bones and teeth. *Nature Protocols*, 2 (7), pp. 1756 -1762.
- Rothe, J. Melisch, C. Powers, N. Geppert, M. Zander, J. Purps, J. Spors, B. Nagy, M.** 2015. Genetic research at a fivefold children's burial from medieval Berlin. *Forensic Sci Int Genet* 15:90–7.
- Sawyer, S. Krause, J. Guschanski, K. Savolainen, V. Pääbo, S.** 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS One* 7:e34131.

Wallin, J.M. Buoncrisiani, M.R. Lazaruk, K.D. Fildes, N. Holt, C.L.

Walsh, P.S. 1998. TWGDAM validation of the AmpFISTR blue PCR amplification kit for forensic casework analysis. *J Forensic Sci* 43:854–70.

Zupanič Pajnič, I. Debska, M. Gornjak Pogorelc, B. Vodopivec

Mohorčič, K. Balažic, J. Zupanc, T. Štefanič, B. Geršak, K. 2016. Highly efficient automated extraction of DNA from old and contemporary skeletal remains. *J Forensic Leg Med* 37:78–86.

Human identification by lice: a next generation sequencing challenge

Pilli E, Agostino A, Vergani D, Salata E, Ciuna I, Berti A, Caramelli D, Lambiase S.

Forensic Sci Int. 2016 Sep;266:e71-e78.

INTRODUZIONE

L'oggetto di studio dell'entomologia e nello specifico dell'entomologia forense sono gli insetti, la loro correlazione con questioni di carattere legale e le applicazioni di questo studio nelle indagini, nei processi penali e nelle controversie civili (*Lord and Stevenson 1986; Catts and Goff 1992*). L'entomologia forense può essere suddivisa in tre branche principali:

- entomologia urbana (*urban entomology*) relativa ai procedimenti giudiziari per infestazioni degli ambienti di lavoro e del vivere civile, nonché consulenze sugli insetti associati alla zootecnia e all'agricoltura;
- entomologia dei prodotti commerciali (*stored products entomology*) relativa ai procedimenti giudiziari che coinvolgono insetti che infestano i beni conservati e/o i loro involucri, in particolare per i prodotti alimentari (entomologia alimentare);
- entomologia medico-legale (*medicocriminal entomology o forensic entomology*) relativa allo studio degli insetti necrofagi coinvolti in eventi di interesse tipicamente medico-legale (omicidi, suicidi, morti improvvise) e altri reati (abuso fisico, negligenze professionali), i quali

contribuiscono al riciclo della materia organica in decadimento cadaverico.

Nonostante questa disciplina abbia origini molto antiche, la sua effettiva applicazione in ambito giudiziario è molto recente. Ad oggi, l'entomologia forense rappresenta uno strumento investigativo indispensabile e complementare a tutte le altre branche della criminalistica poiché, dal preciso campionamento e dalla identificazione degli artropodi che si trovano sulla scena *criminis*, è possibile determinare una correlazione con i parametri ambientali specifici e stimare (Magni, Massimelli and Messina 2008; Di Luise, Magni and Sarano, 2007):

- l'intervallo *post mortem* (PMI) attraverso la conoscenza delle specie di insetti e del loro ciclo di sviluppo;
- il luogo del delitto ed eventuali spostamenti del cadavere attraverso la conoscenza delle specie endemiche di luoghi diversi;
 - le eventuali lesioni che potrebbero essere la causa della morte;
- l'eventuale consumo di sostanze stupefacenti o l'esposizione a sostanze tossiche.

Un ulteriore supporto allo sviluppo delle conoscenze e delle applicazioni dell'entomologia giunge, negli ultimi anni, dal rapido e progressivo avanzamento delle frontiere della biologia molecolare moderna che hanno rivoluzionato non solo il mondo della medicina, della

microbiologia e delle biotecnologie ma anche il settore forense, dove l'analisi del DNA a scopi identificativi, ha subito una eccezionale propulsione, soddisfacendo la richiesta di metodi affidabili, sensibili e altamente discriminanti. La biologia molecolare contribuisce pertanto attivamente alle analisi di entomologia forense intraprendendo due diverse linee di applicazione specifica. La prima applicazione, prettamente sistematica, prevede l'identificazione della specie dell'insetto mediante analisi di marcatori molecolari specifici, cioè di regioni del genoma comuni agli appartenenti a una stessa specie e/o a specie diverse, così da consentire di classificare o identificare un individuo di una specie piuttosto che di un'altra. Questa tecnica può così fornire supporto all'approccio classico di classificazione, basato sull'individuazione di caratteri morfologici attraverso le cosiddette "chiavi dicotomiche", superandone il limite connesso, talvolta, all'enorme difficoltà o impossibilità di distinguere uova, pupe e taluni stadi larvali dagli stadi analoghi di specie diverse. Ciò consente quindi una discriminazione delle specie su base genotipica in supporto a quella su base fenotipica (Amendt *et al.* 2004). Recenti studi sono stati pubblicati sull'identificazione per via biomolecolare degli invertebrati di interesse medico-legale (Gillariourtua, Salona Bordas *et al.* 2013; Boehme, Amendt, and Zehner 2012). Non solo le larve e gli adulti campionati sul luogo del reato ma anche i pupari cioè gli involucri protettivi della larva in metamorfosi si sono dimostrati utili ai fini identificativi (Mazzanti *et al.* 2010). Inoltre, in ambito forense, lo studio degli insetti potrebbe essere di notevole importanza al fine di individuare miasi o infezioni nosocomiali (M. Jang *et al.* 2013; G. Mowlavi *et al.*

2011) in cui gli insetti possono agire da vettori di malattie evidenziando lo stato di negligenza in cui sono costrette a vivere alcune persone anziane. La seconda applicazione è invece l'individuazione del materiale genetico umano presente nei reperti entomologici. La finalità dal punto di vista forense, a tal proposito, è di concorrere all'identificazione del soggetto e all'accertamento dell'esistenza di una correlazione tra insetto, luogo del reato e vittima. Gli insetti necrobionti che si nutrono dei tessuti del corpo in decomposizione e gli insetti ematofagi che succhiano il sangue dai corpi mantengono indigesto al loro interno il materiale biologico della carcassa per un tempo variabile a seconda dell'esemplare. Il materiale indigesto costituisce così una fonte di DNA utilizzabile ai fini investigativi, fornendo indicazioni sulle relazioni spazio temporali degli eventi che hanno avuto luogo sulla scena del crimine e permettendo di associare una vittima ad un determinato luogo. Qualora vi fosse uno spostamento/occultamento del cadavere, il ritrovamento e la localizzazione di particolari insetti, o meglio del loro contenuto di DNA umano, potrebbe rappresentare la prova del passaggio della vittima in un determinato luogo, avvalorando così la determinazione del PMI. Similmente, tale approccio potrebbe chiarire, in base allo studio delle successioni, dove e/o su cosa si siano sviluppate le larve qualora la "fonte alimentare" delle stesse sia dubbia (Wells et al. 2001; Zehner, Amendt and Krettek 2004; Spitaleri et al. 2006). Un ulteriore impiego potrebbe essere quello di supportare le indagini nell'ambito delle violenze sessuali, sfruttando la capacità delle larve di svilupparsi a livello dei genitali e di nutrirsi dello sperma (Clery 2001). In merito all'analisi del materiale

genetico umano ottenuto da reperti di natura entomologica, sono stati condotti diversi studi che hanno sperimentato l'estrazione e il successivo sequenziamento del DNA umano, utilizzando matrici proprie dell'insetto quali per esempio:

- detriti e feci, che possono contenere quantità di DNA umano sufficienti all'identificazione del cadavere a scapito del quale essi si sono nutriti (*Replogle et al., 1994*);
- il contenuto gastrico o *crop*, ovvero l'appendice esofagea adibita all'immagazzinamento del cibo.

In diversi studi è stata dimostrata la possibilità di analizzare sia il DNA mitocondriale sia i marcatori *STR* nucleari da questa matrice, prelevata dalle larve sviluppatesi sui resti cadaverici (*Wells et al. 2001; Linville et al. 2004; Zehner et al. 2004; Li et al. 2011; Chavez-Briones et al. 2013*) o dai coleotteri sviluppatesi su ossa umane (*DiZinno et al. 2002*). Recentemente un gruppo di studiosi ha dimostrato la possibilità di rilevare l'antigene prostatico specifico (PSA) e di tipizzare i marcatori del cromosoma Y partendo dall'estratto del "gozzo" di ditteri allevati parecchie ore prima (fino a 145 ore) su tessuti ricoperti di sperma umano (*Clery 2001*), aprendo così la strada all'applicazione nei casi di violenza sessuale - omicidi, in cui i cadaveri vengono rinvenuti molti giorni dopo l'evento delittuoso. In tali contesti infatti, spesso le vittime sono ritrovate in condizioni tali da non permettere l'isolamento delle tradizionali tracce biologiche dell'aggressore sul corpo e sui vestiti. Pertanto sfruttare la peculiarità dell'entomofauna di colonizzare velocemente gli orifizi del

cadavere potrebbe essere utile ai fini di identificare il colpevole del reato. In casi di violenza sessuale è anche possibile il trasferimento occasionale dall'assalitore alla vittima di pidocchi o piattole, parassiti comuni dell'uomo che assumono sangue umano e che potrebbero essere utilizzati per accertare la correlazione fisica dell'ospite originario con la vittima, considerato che è stata verificata la rilevabilità del DNA mitocondriale umano nei loro resti (*Lord et al. 1998; Mumcuoglu et al. 2004*).

SCOPO DEL LAVORO

Il presente lavoro è consistito nell'investigare la possibilità di sfruttare *Pediculus humanus capitis* come substrato biologico utile ai fini dell'identificazione personale umana. A tale scopo, per la prima volta in questo lavoro, la tecnologia *NGS* è stata utilizzata per analizzare il DNA nucleare umano, degradato ed in scarsa quantità, estratto dal tratto gastrointestinale degli insetti, rinvenuti, in sede di sopralluogo, nell'abitazione di una anziana Signora deceduta, in ospedale, inseguito ad una parassitosi. Giunta in ospedale, la Signora si presentava in uno stato di profondo abbandono e scarsa igiene, pertanto, al fine di stabilire le cause della morte il Giudice ha disposto l'autopsia e l'ispezione dell'abitazione. Era infatti necessario appurare se lo stato di abbandono e scarsa igiene in cui versava la Signora dipendessero dall'incuranza dei familiari. Durante il sopralluogo nessuna traccia biologica di interesse investigativo è stata riscontrata in casa, solo diversi pidocchi sono stati rinvenuti nella stanza in cui la signora dormiva ed in quella degli ospiti dove la figlia diceva di dormire lei o la badante. Pertanto, data l'assenza di evidenti segni della presenza di altre persone nella casa e la grande quantità di pidocchi riscontrata in casa, si è ipotizzato che qualsiasi altra persona avesse vissuto con la Signora, sarebbe stata a sua volta infestata dagli insetti. Pertanto, i pidocchi rinvenuti nella camera degli ospiti, dove la figlia diceva di vivere o lei o la badante, sono stati campionati al fine di stabilire se gli insetti potessero aver infestato la figlia e/o la badante.

MATERIALI E METODI

1. Il campionamento

In sede di sopralluogo sono stati repertati dall'entomologo circa una trentina di esemplari di insetti, riconosciuti su base morfologica come *Pediculus humanus capitis* e conservati all'interno di falcon a temperatura ambiente per le analisi genetiche. Ulteriori altri individui sono stati campionati e conservati a temperatura ambiente in EtOH 75% per le indagini entomologiche di identificazione della specie di appartenenza. *Pediculus humanus capitis* appartiene alla famiglia Pediculidae, dell'ordine Anoplura, che include degli organismi ectoparassiti permanenti ed obbligati, comunemente noti come pidocchi o piattole, i quali parassitano l'uomo, prediligendo le aree del corpo riccamente ricoperte di formazioni pilifere come la testa, il pube, oppure le fibre degli indumenti, insediandosi nelle aree dove gli abiti sono particolarmente aderenti al corpo. Questi insetti si nutrono del sangue dell'ospite introducendolo nel sacco trofico del loro apparato digerente, immagazzinando così anche il materiale genetico dell'ospite. I processi di digestione enzimatica tuttavia sono responsabili della degradazione del materiale biologico introdotto quindi, a sua volta anche del DNA.

Le indagini entomologiche sono state effettuate presso la sezione di Entomologia Forense del Dipartimento di Sanità pubblica Medicina Sperimentale e Forense dell'Università di Pavia mentre le analisi genetiche sono state effettuate presso i Laboratori di Antropologia

Molecolare dell'Università di Firenze in collaborazione con il Reparto Investigazioni Scientifiche dell'Arma dei Carabinieri di Roma.

2. Ispezione microscopica

Al fine di investigare le caratteristiche morfostrutturali degli insetti e confermare la specie di appartenenza degli stessi, tutti i campioni sono stati esaminati allo stereomicroscopio (Leica MZ12 –*Leica Microsystems* srl, Milano). L'indagine entomologica è stata condotta in accordo con la metodologia di identificazione classica attraverso le chiavi dicotomiche.

3. Le analisi genetiche

Dopo un'accurata ispezione allo stereomicroscopio, gli esemplari di *Pediculus humanus capitis* conservati in falcon sterili sono stati usati per l'analisi genetica. I campioni sono stati pertanto sottoposti ad una rapida procedura di pulizia dai contaminanti esterni con lavaggi sequenziali in varechina (soluzione 5%) per 30 secondi, etanolo e acqua ed infine sottoposti alla rimozione delle appendici addominali e del capo. Dopo averli fatti asciugare, a circa 20 esemplari è stata praticata un'incisione longitudinale a livello dell'esoscheletro dell'addome al fine di facilitare il processo di lisi cellulare e di successiva estrazione del DNA ed immersi direttamente in circa 1 ml di buffer *dell'Hexagon OBTI test* in due differenti *eppendorf* (10 in un'*eppendorf*, 10 in un'altra). Prima di sottoporre infatti i campioni alla procedura di estrazione è stato eseguito il test confermativo, *Hexagon OBTI*, per determinare la natura della traccia contenuta all'interno del tratto gastrointestinale degli insetti. Tale test, specifico per il sangue umano, sfrutta una reazione

immunocromatografica. Tra gli antigeni presenti nel sangue umano, cioè l'emoglobina, e gli anticorpi monoclonali anti-hHb mobili legati a particelle colorate e presenti nella striscia *test* si crea un complesso antigene-anticorpo. L'immunocomplesso così formatosi migra verso la zona *test* dove viene catturato da un secondo anticorpo immobilizzato diretto contro l'emoglobina e precipita formando una linea blu di *test*: risultato positivo. Tuttavia al fine di verificare il corretto funzionamento del *test* è necessaria la formazione di una seconda linea, detta di controllo, che si genera dal legame tra i componenti che non hanno formato l'immunocomplesso ed anticorpi IgG immobilizzati. Quando la "test bar" presenta una sola banda in corrispondenza della zona di controllo, il *test* è negativo, non siamo in presenza di sangue umano o la sua concentrazione è al di sotto della soglia di rilevabilità del *test*. Il *test* che invece non presenta alcuna banda indica che c'è stato un problema di funzionamento del *test* e pertanto quest'ultimo deve essere ripetuto. Nel caso in cui invece si ottengano due bande, il *test* è positivo: il *test* ha funzionato e siamo in presenza di una traccia ematica umana.

Dopo circa 30 minuti di incubazione nel *buffer* del *kit*, due gocce di soluzione vengono trasferite nell'apposito dispositivo del *test* ed i risultati vengono valutati dopo circa 10 minuti.

3.1 Estrazione del DNA

Come suggerito da *Johnston et al.* (*Johnston et al. 2008*), le aliquote di *buffer* OBTI e campione sono state sottoposte ad estrazione previa concentrazione con Amicon Ultra-0.5ml *centrifugal filters* (Millipore).

L'estrazione è stata effettuata mediante *QIAmp DNA Investigator kit* seguendo il protocollo per estrazione di DNA da tessuto (*Isolation of total DNA from tissue*). Il campione di confronto rappresentato da materiale ematico è stato sottoposto ad estrazione del DNA tramite *Chelex® 100*.

3.2 Quantificazione del DNA

Una volta estratto, il DNA è stato quantificato mediante l'utilizzo di due *kits*, il *Quantifiler® Duo* ed il *Quantifiler® Trio* (*Life technologies*). Questi permettono di quantificare, simultaneamente in uno stesso campione, la quantità totale di DNA umano e l'eventuale presenza di DNA maschile. La quantificazione con il *Quantifiler® Duo* viene fatta tramite due *targets* uno di 140 bp specifico per il DNA autosomico umano e l'altro di 130 bp specifico per il DNA umano maschile.

I *targets* utilizzati dal *Quantifiler® Trio* sono invece tre: *small autosomal*, *large autosomal* ed *Y target*.

- Lo *small autosomal* è di piccole dimensioni, circa 80 bp, e permette la quantificazione anche di campioni altamente degradati;
- Il *large autosomal* (214 bp) viene utilizzato principalmente come indicatore della degradazione del DNA del campione. Dal rapporto tra lo *small autosomal* ed il *large autosomal* è possibile calcolare l'indice di degradazione di un campione. L'indice di degradazione risulta un indicatore sensibile di quanto il DNA che stiamo analizzando sia degradato. I valori che si possono ottenere sono: 1. < 1, il DNA non è degradato; 2. da 1 a 10, il DNA è da

leggermente a moderatamente degradato e 3. > 10 o nessun valore, il DNA è notevolmente degradato.

- *Target* cromosoma Y(75 bp) consente la quantificazione del DNA umano maschile ed è particolarmente utile quando si deve quantificare la presenza del DNA maschile in una miscela maschio-femmina.

3.3 SNPs profiling e preparazione librerie

Una volta quantificati i campioni, è stata allestita una libreria genomica secondo il protocollo *Ion AmpliSeq™ Library Preparation for Human Identification Applications* in combinazione con *HID-Ion AmpliSeq™ Panels* al fine di amplificare le regioni *target* di interesse a partire dal DNA genomico, in accordo con il manuale d'uso. Questo pannello è composto da 30 *SNPs* del cromosoma Y e 90 *SNPs* autosomici selezionati da *Pakstis et al.* e da *ID Consortium (Phillips et al. 2007; Karafet et al. 2008)*.

I campioni sono stati amplificati in duplicato. Gli ampliconi ottenuti sono stati trattati con il reagente FuPa, costituito da una miscela di enzimi che preparano l'amplicone per la ligazione con gli adattatori, ovvero digeriscono le estremità e fosforilano l'amplicone. Successivamente ad entrambe le estremità è stato legato l'adattatore e ad una delle due estremità il *barcode* che ha permesso di etichettare il campione. Al termine di questo processo si è proceduto con la purificazione mediante *AMPure® XP Reagents (Agencourt)* e successiva quantificazione mediante *Ion Library Quantitation kit (Thermo Fisher Scientific)*. Infine

un *pool* 20 pM è stato creato riunendo differenti librerie da usare per la PCR emulsio-clonale.

3.3.1 PCR emulsio-clonale e sequenziamento

Durante questa fase analitica le singole molecole di DNA della libreria sono state denaturate e ibridate a singole *Ion PGM™ Template OT2 200 Ion Sphere™ Particles (ISP)*. Ad ogni nanosfera è stata legata una singola molecola di DNA per evitare di avere al termine della reazione di PCR sfere policlonali. La PCR emulsio-clonale è stata eseguita mediante lo *Ion OneTouch™ 200 Template Kit v2 (Thermo Fisher Scientific)* sul *OneTouch™ 2 (OT2) instrument (Thermo Fisher Scientific)*. Introducendo i reagenti di PCR, l'olio minerale e i *primers* biotinilati nel filtro sotto pressione è stata creata un'emulsione. All'interno delle micelle di acqua-olio sulle singole superfici delle sfere è stata effettuata una PCR fino a generare una genoteca di quello specifico frammento legato alla singola sfera. L'emulsione è stata rotta e creato un *pellet* di ISP. Successivamente i prodotti di PCR sono stati arricchiti mediante lo *Ion OneTouch™ Enrichment System (Thermo Fisher Scientific)* utilizzando lo *Ion PGM™ Template OT2 200 Kit (Thermo Fisher Scientific)*. Il sequenziamento è stato quindi allestito utilizzando *Ion PGM™ Sequencing 200 Kit v2* ed il *316 Ion™ Chips* seguendo le raccomandazioni del manuale.

4. Analisi biostatistica

Inizialmente l'analisi dei dati è stata effettuata utilizzando *Torrent Server v4.2*. L'analisi della copertura è stata effettuata mediante lo *Ion Torrent*

Coverage Analysis Plug-in v4.2. I dati sono stati quindi esportati dal *plugin* e le successive analisi sono state effettuate con Excel[®] (Microsoft) e Minitab[®]17 (Minitab inc. USA). Una volta ottenuto il profilo dal campione oggetto di indagine e da quello di confronto si è proceduto all'osservazione dei profili ed alla comparazione degli stessi. Quando non viene evidenziata alcuna sovrapposibilità tra i profili in analisi si può parlare di *no match*, ovvero di esclusione: i profili confrontati non sono compatibili e quindi la traccia reperita sulla scena del crimine non appartiene al campione di confronto indagato; viceversa se è presente una sovrapposibilità completa per tutti i *loci* analizzati si parla di *match*. Una volta osservata la compatibilità genetica fra due profili viene effettuata una valutazione biostatistica dei risultati, così da poter affiancare alla compatibilità dei dati genetici un valore numerico a supporto della compatibilità osservata. A questo proposito, al fine di poter effettuare il calcolo della *Random match Probability* (RMP), ovvero della probabilità di condivisione casuale, che esprime la possibilità di selezionare il genotipo osservato in una popolazione di individui *random* tra loro non correlati, è stata utilizzata un'applicazione di calcolo *online* denominata *FROG-kb* (*Forensic Research/Reference on Genetics-knowledge base*, <http://frog.med.yale.edu/FrogKB/>). Tale applicazione si avvale di un *database* di riferimento denominato *ALFRED* (*Allele FREquency Database*) che raccoglie tutte le frequenze alleliche di ciascuno *SNPs* in ciascuna popolazione e sfrutta per il calcolo l'equilibrio di *Hardy-Weinberg*, secondo il quale la relazione tra le frequenze alleliche e

genotipiche, da una generazione a quella successiva, sono costanti nel tempo, come espresso nella formula:

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

Per ciascun *locus* la frequenza degli alleli viene moltiplicata come $2pq$ se vi è una condizione di eterozigosi per quel marcatore, viceversa viene moltiplicata per se stessa, cioè elevata al quadrato se c'è una condizione di omozigosi. Una volta ottenuta la frequenza genotipica specifica di ciascun marcatore, l'applicazione calcola la *RMP* moltiplicando la frequenza genotipica di ciascun *locus* secondo la regola del prodotto (Pakstis et al. 2010; Kidd et al. 2012; Sánchez et al. 2006). Nell'indagine forense l'espressione statistica più usata è tuttavia il rapporto di verosimiglianza o *likelihood ratio* (*LR*) che rappresenta il rapporto delle probabilità di due ipotesi alternative, cioè l'ipotesi di identità contro quella di condivisione casuale.

$$LR = \frac{H_P}{H_D}$$

dove H_p è detta ipotesi dell'accusa perché sostiene che la traccia appartenga al sospettato e H_d è detta ipotesi della difesa perché sostiene che la traccia rinvenuta possa provenire da un individuo diverso dal sospettato. Concretamente *LR* è dato dall'inverso della probabilità di condivisione casuale (*RPM*). Secondo questo calcolo, più è elevato il

rapporto di verosimiglianza, maggiore è la forza con cui si asserisce l'identità.

RISULTATI E DISCUSSIONE

In seguito ad un'accurata indagine microscopica effettuata tramite il metodo di classificazione classico delle chiavi dicotomiche, è stato possibile confermare che tutti i campioni raccolti in sede di sopralluogo appartenevano alla stessa specie, *Pediculus humanus capitis*. L'ispezione microscopica ha permesso inoltre di mettere in evidenza che i diversi individui campionati avevano dimensioni dissimili a causa presumibilmente del diverso stadio di crescita in cui si trovavano e alcuni insetti presentavano una diversa colorazione dell'addome (scuro o trasparente) rispetto agli altri (Figura 1).

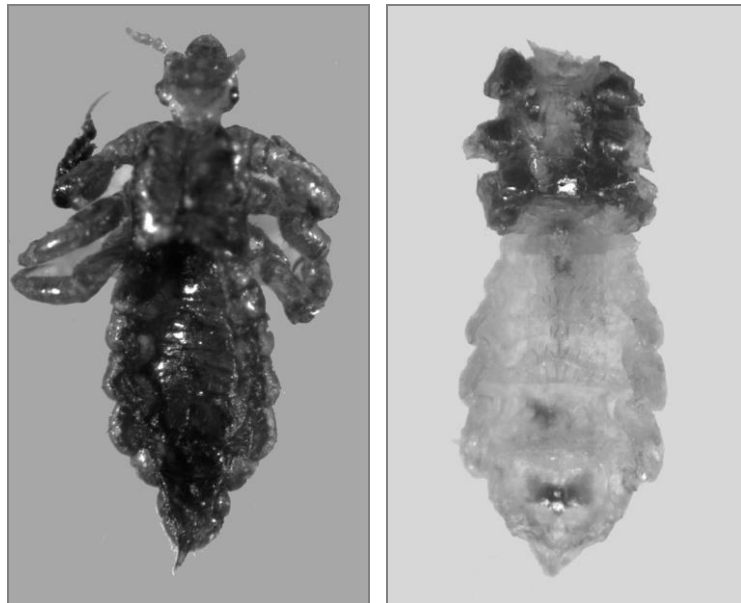


Figura 1: Esempio di un individuo che presentava l'addome scuro (a sinistra) ed esempio di un individuo trasparente (a destra).

La differente colorazione dell'addome, presumibilmente dovuta alla muta, potrebbe però essere associata anche ad una diversa fase di digestione in cui si trovavano gli insetti al momento della loro morte. Dieci pidocchi che presentavano una colorazione scura dell'addome e dieci "trasparenti" sono stati testati per la presenza di sangue umano mediante l'*Hexagon OBTI test*. Il risultato, ottenuto dopo una trentina di minuti di incubazione e dieci di reazione, è stato positivo per gli insetti scuri e negativo per quelli trasparenti. Al fine di valutare se l'incisione dorsale possa aver facilitato la fuoriuscita di materiale biologico dal tratto gastrointestinale degli insetti come suggerito da *Lord et al.* 1998, oltre ai dieci insetti scuri precedentemente sezionati (*Pediculus A*), altri dieci insetti scuri privi di sezione dorsale (*Pediculus B*) sono stati sottoposti ad estrazione. Successivamente la concentrazione del DNA estratto è stata valutata mediante il kit *Quantifiler*[®] *Duo*. Nessun risultato è stato ottenuto sia dal *target* autosomico che dal *target Y* per il campione *Pediculus B* e *Bbis* (due repliche dello stesso estratto) e solo 3 pg/μl dal *target* autosomico del campione *Pediculus A* e *Abis*. Dal momento che tale valore è risultato essere al di sotto del limite di rilevazione (LOD *limit od detection*) del kit, abbiamo deciso di effettuare una nuova quantificazione sfruttando il nuovo kit *Quantifiler*[®] *Trio* che ha un LOD di 0.8 pg/ul come riportato nel manuale del kit. Il nuovo kit che utilizza *loci target* in multi-copia, è stato ottimizzato rispetto al precedente Duo sia in termini di sensibilità che di rilevazione di campioni degradati. I risultati di quantificazione sono riportati in Tabella 1.

QUANTIFILER TRIO KIT				
Campioni	Quantità (ng/μl)			ID
	LAT	SAT	Y Target	
A	0.0069	0.0128	0.00	1.85
A bis	0.0072	0.0073	0.00	1.01
B	0.00	0.00	0.00	
B bis	0.00	0.00	0.00	
ENC	0.00	0.00	0.00	
ENC	0.00	0.00	0.00	

Tabella 1: Quantificazione dei campioni estratti mediante *Quantifier Trio DNA Quantification kit*. ECN: controllo negativo di estrazione. L'indice di degradazione (ID) è stato calcolato dall'*HID Real-Time PCR software* come il rapporto fra la concentrazione del DNA (ng/μl) fra lo *small autosomal target* (SAT) e la concentrazione del DNA (ng/μl) del *large autosomal target* (LAT). A = *Pediculus A*; Abis = *Pediculus Abis*; RS = *Reference sample*.

Come si può osservare dalla Tabella 1: nessuna contaminazione si è verificata nel controllo negativo di estrazione; nessun risultato è stato ottenuto dall'estratto del *Pediculus B* e *Bbis* sia per lo *small* che per il *large autosomal* che per l'*Y target*; valori invece di quantificazione simili, pari a 6.9 e 7.2 pg/ul sono stati ottenuti per il *large autosomal target* del campione *Pediculus A* ed *Abis* rispettivamente e di 12.8 e 7.3 pg/ul per lo *small autosomal target*. La differenza che è stata osservata nei valori di quantificazione dello *small autosomal target* nelle due repliche effettuate del campione *Pediculus A* ed *Abis* potrebbero essere dovute alla variabilità casuale di pipettata associata al diverso grado di

degradazione del contenuto gastro-intestinale al momento in cui gli insetti sono stati raccolti ed analizzati. In Tabella 1 si può inoltre osservare come nessuna quantificazione è stata effettuata per il *target* Y. Tale risultato negativo ottenuto dal marcatore Y ci ha portato a supporre che il DNA estratto dagli insetti potesse appartenere ad un individuo di sesso femminile. I diversi risultati di quantificazione che si sono osservati per il *Pediculus* A e B possono essere spiegati con l'assenza di incisione dorsale nei *Pediculus* B che sembra essere fondamentale per permettere la fuoriuscita del materiale biologico dal tratto gastro-intestinale. L'indice di degradazione, riportato in Tabella 1, è risultato pari a 1.85 e 1.01 per *Pediculus* A e B rispettivamente. Come riportato nel manuale del *kit* di quantificazione, valori di indice di degradazione compresi fra 1 e 10 indicano una moderata degradazione del materiale genetico quantificato. Dopo aver effettuato la quantificazione, gli estratti di *Pediculus* A e B sono stati amplificati mediante *AmpFlSTR® NGM SElect™ PCR Amplification Kit* (Thermo Fisher Scientific, Oyster Point, CA). Dal momento che un profilo parziale/degradato è stato ottenuto dal *Pediculus* A e nessun profilo dal *Pediculus* B tramite la metodologia *STR* classica, abbiamo deciso di impiegare la nuova tecnologia *NGS* su tale tipologia di campioni.

Successivamente anche il campione di riferimento della vittima (sangue) prelevato al momento dell'autopsia è stato estratto e quantificato. I valori di quantificazione sono risultati pari a 5.49 e 2.48 ng/ul rispettivamente per il *large* e lo *small autosomal target*. L'indice di degradazione

calcolato è risultato pari a 0.45 valore questo che evidenzia, come atteso, l'assenza di degradazione.

In seguito alla quantificazione, al fine di allestire la libreria, gli estratti sia dei campioni di insetto che del campione di riferimento sono stati amplificati mediante *HID-Ion AmpliSeq™ Identity Panel* utilizzando un protocollo che prevedesse differenti cicli di amplificazione a causa dei diversi risultati di quantificazione ottenuti per i campioni *Pediculus A* e per il campione di riferimento. In seguito, un *pool* equimolare 20 pM è stato preparato ed i campioni sono stati fatti correre su sequenziatore *Thermo-Fisher Scientific*. I risultati della corsa (copertura e parametri di qualità) sono riportati in Tabella 2.

Sample	Bases	≥Q20	Reads	Total SNP Coverage	Mean Read Length
A	5.579.081	4.883.061	59.205	24.903	94bp
Abis	6.721.047	5.808.011	72.681	22.750	92bp
RS	5.910.513	5.491.694	72.917	66.864	81bp

Tabella 2: Risultati della corsa NGS. A= *Pediculus A*; Abis= *Pediculus Abis*; RS= *Reference sample*.

Come si evince dalla Tabella 2 la copertura totale degli *SNPs target* è risultata bassa rispetto al numero totale delle *reads* ottenute. In particolare i risultati ottenuti da *Pediculus A* e *Abis* hanno mostrato che la maggior parte delle *reads* non coprivano la regione degli *SNPs* di interesse suggerendo questo che non tutti gli *SNPs* di interesse venivano rappresentati da un numero significativo di *reads* (molti *SNPs*

presentavano infatti un *coverage* sotto 30X) a causa presumibilmente della degradazione delle regioni *target*. Pertanto, la mancata copertura di questi *SNPs* può aver causato una riduzione nel numero totale delle *reads* che hanno contribuito al *coverage*.

Come è possibile osservare in Figura 2 (A e B) *Pediculus A* e *Pediculus Abis* presentavano un profilo di copertura simile.

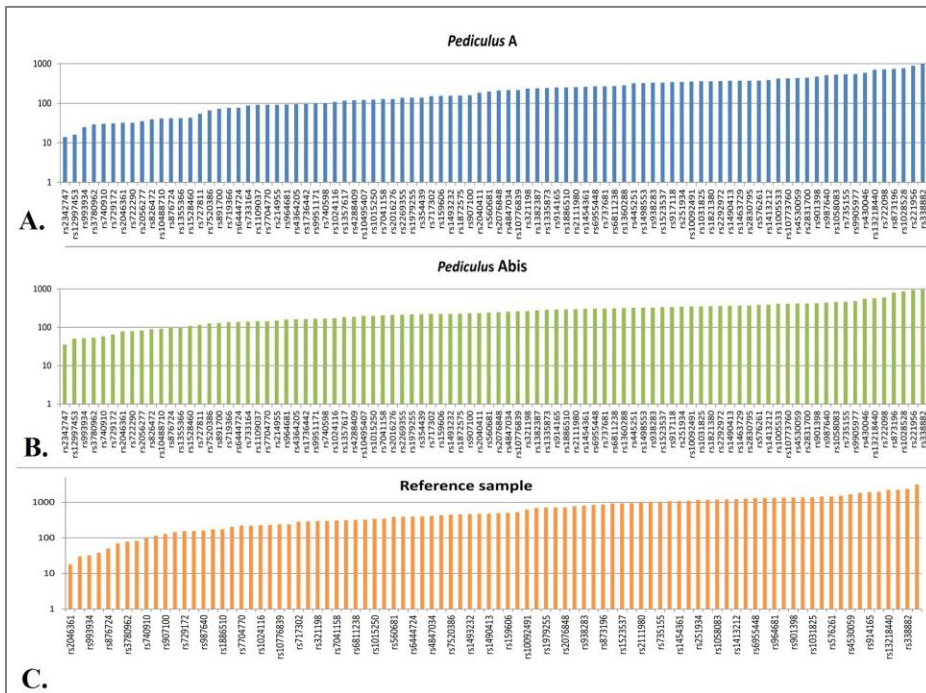


Figura 2: Copertura del *Pediculus A/Abis* e del campione di riferimento. Gli *SNPs* sono organizzati in ordine decrescente di copertura. Sull’asse delle ascisse sono riportati gli *SNPs* e su quello delle ordinate la copertura.

Il confronto fra il profilo di copertura di *Pediculus A/Abis* e quello del campione di riferimento (Figura 2 A/B e C) ha evidenziato alcune

differenze presumibilmente dovute ad uno sbilanciamento nella creazione del *pool*. Se gli *SNPs* analizzati vengono organizzati in ordine crescente rispetto al peso molecolare, il confronto fra il profilo dei campioni (Figura 3 D ed E) e quello del campione di confronto (Figura 3 F) ha evidenziato una copertura non omogenea degli *SNPs* dei campioni di *Pediculus A* e *Abis* rispetto a quelli del campione di riferimento.

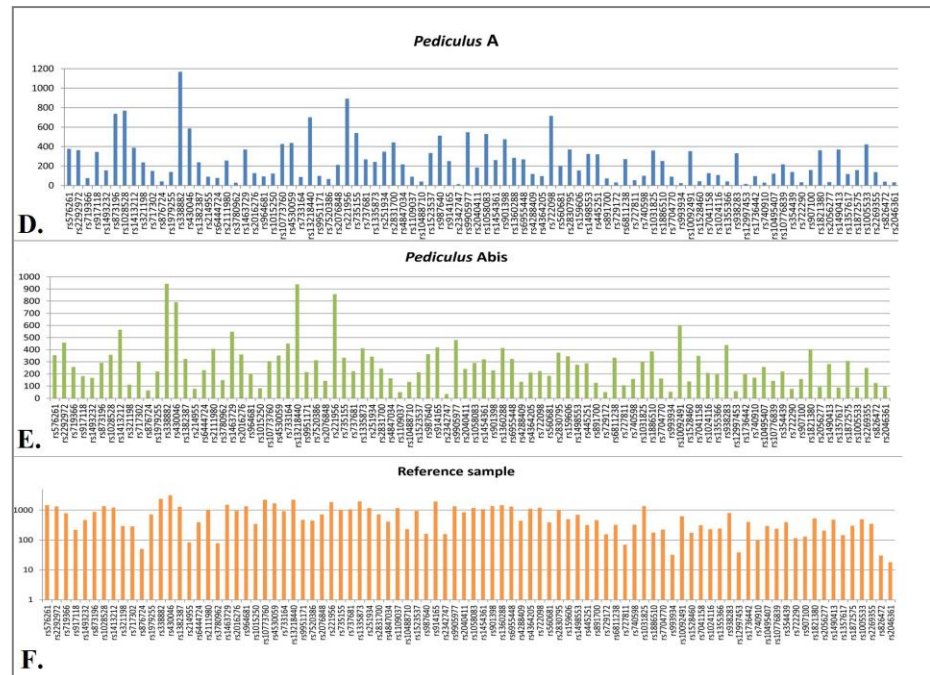


Figura 3: Copertura del *Pediculus A/Abis* e del campione di riferimento. Gli *SNPs* sono organizzati in base alla dimensione, dal più piccolo al più grande. Sull’asse delle ascisse sono riportati gli *SNPs* e su quello delle ordinate la copertura.

La differente altezza dei picchi che si è potuta osservare fra i profili di *Pediculus A* e *Abis* può presumibilmente essere ricondotta alla

degradazione del campione, come precedentemente descritto da Salata et al. (Salata et al. 2015).

Il profilo genetico dei campioni di insetto e del campione di riferimento ottenuti in seguito all’analisi con *HID SNP Genotyper Plug-in v4.2* è riportato in Figura 4.

		SNPs																																			
Samples		rs1498413	rs720386	rs417034	rs560681	rs1095407	rs891700	rs1413212	rs876724	rs1109637	rs993934	rs12997453	rs907100	rs1357617	rs1872275	rs4364205	rs1872275	rs6444724	rs2946361	rs6811238	rs1979255	rs717302	rs159806	rs7704770	rs231954	rs338882	rs13218440	rs214955	rs72811	rs6955448	rs971718	rs321198	rs737681	rs1002491	rs438849		
A		A	G	A	A	A	C	T	G	A	N	Z	C	T	G	G	C	C	A	G	G	A	A	A	A	A	A	A	A	C	C	T	C	C	A	C	
A bis		A	G	A	A	A	C	T	G	A	A	A	C	T	G	G	C	C	A	G	G	A	A	A	A	A	A	A	A	C	C	T	C	C	A	C	
Ref		A	G	A	A	A	C	T	G	A	A	A	C	T	G	G	C	C	A	G	G	A	A	A	A	A	A	A	A	C	C	T	C	C	A	C	
		SNPs																																			
Samples		rs206277	rs1016250	rs7041158	rs1463729	rs1480286	rs10770839	rs625472	rs734155	rs789062	rs746598	rs964681	rs1408553	rs901398	rs10488710	rs2079848	rs2269355	rs2111980	rs1335873	rs1077390	rs1335873	rs1886510	rs1058083	rs554439	rs1454361	rs72290	rs873196	rs4530659	rs2016276	rs1821380	rs1528460	rs2422747	rs729172	rs453046	rs1382387	rs9965977	
A		C	G	C	T	C	G	T	C	A	A	C	C	C	C	A	C	T	A	A	A	G	G	A	A	G	C	T	C	C	C	T	G	C	A	A	
A bis		C	G	C	T	C	G	T	C	A	A	C	C	C	C	A	C	T	A	A	A	G	G	A	A	G	C	T	C	C	C	T	G	C	A	A	
Ref		C	G	C	T	C	G	T	C	A	A	C	C	C	C	A	C	T	A	A	A	G	G	A	A	G	C	T	C	C	C	T	G	C	A	A	
		SNPs																																			
Samples		rs49910	rs938383	rs292972	rs149332	rs9951171	rs1024116	rs719366	rs576261	rs1031625	rs445251	rs1005533	rs1523337	rs722098	rs2830795	rs2931700	rs914165	rs221956	rs733164	rs987640	rs2940411	rs1028238															
A		A	T	C	A	A	C	C	A	C	C	C	C	A	A	A	G	G	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
A bis		A	T	C	A	A	C	C	A	C	C	C	C	A	A	A	G	G	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Ref		A	T	C	A	A	C	C	A	C	C	C	C	A	A	A	G	G	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Figura 4: Profili ottenuti dall’analisi degli SNPs utilizzando la piattaforma *IonTorrent PGM*. La completa sovrapposizione osservata fra i campioni ed il riferimento è mostrata in grigio; gli effetti stocastici di *droup-out* e *drop-in* sono riportati in chiaro ed infine in bianco è riportata l’assenza di risultato. A=*Pediculus A*; Abis= *Pediculus Abis* e Ref= campione di confronto.

Come è possibile osservare in Figura 4, dal campione di confronto è stato ottenuto un profilo genetico completo. Da *Pediculus A* è stato ottenuto un profilo genetico completo ad eccezione dello SNP rs12997453 che non è stato identificato e dello SNP rs1454361 che risulta essere affetto da un

singolo evento di *drop-in* che ha cambiato il genotipo da AA (genotipo atteso) ad AT (genotipo osservato: A coperto 258X e T 36X). La copertura dello *SNP* rs12997453 è 58 ma l'estremo sbilanciamento che è stato osservato fra la copertura dei due alleli (A 49 e G 9), non ha permesso di identificare il *locus*. Anche da *Pediculus Abis* è stato ottenuto un profilo completo ad eccezione di un singolo evento di *drop-out* osservato allo *SNP* rs1024116 che ha cambiato il genotipo atteso CT a quello osservato CC. Successivamente, il bilanciamento degli alleli eterozigoti è stato valutato ed i risultati sono riportati in Figura 5.

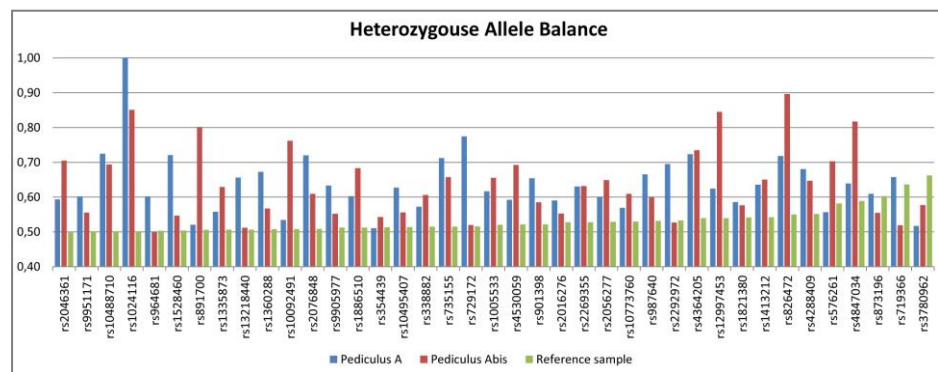


Figura 5: Bilanciamento degli alleli eterozigoti in *Pediculus A*, *Abis* e nel campione di riferimento. Sull'asse delle ascisse sono riportati gli *SNPs* analizzati e sull'asse delle ordinate il bilanciamento allelico espresso in percentuale.

Come si può osservare dai risultati riportati in Figura 5, il campione di riferimento presentava un profilo eterozigote ben bilanciato per tutti gli *SNPs* analizzati. I profili dei campioni di insetto hanno mostrato, invece, dei profili molto sbilanciati, presumibilmente a causa degli effetti stocastici che possono intervenire quando scarse quantità di DNA di

partenza vengono amplificate. Dalla comparazione fra il profilo genetico ottenuto da *Pediculus* A ed Abis è stato possibile osservare una completa sovrapposizione di tutti gli *SNPs* analizzati, ad eccezione dei *loci* in cui sono stati osservati gli effetti stocastici precedentemente descritti. Anche la comparazione fra il profilo ottenuto dal campione di confronto e quello dei campioni di insetto ha evidenziato una completa compatibilità per tutti gli *SNPs* analizzati. Il calcolo dell'*RMP*, effettuato sfruttando le frequenze relative alla popolazione Europea dell'applicazione *FROG-kb*, è risultato essere pari a $1,82 \times 10^{-38}$. Il valore di *LR* stimato del profilo *SNPs* ha evidenziato che era $5,49 \times 10^{37}$ volte più probabile che il profilo dei campioni di insetto provenisse dalla vittima piuttosto che da una persona sconosciuta appartenente alla popolazione europea. Pertanto, le evidenze genetiche hanno fornito un supporto estremamente forte all'ipotesi che il profilo ottenuto dall'analisi del DNA estratto dal tratto gastrointestinale dei campioni in analisi appartenesse al campione di confronto (vittima).

CONCLUSIONI

Le tecniche di tipizzazione molecolare forniscono un valido supporto alle indagini entomologiche forensi sia per l'identificazione di specie che per l'identificazione di materiale genetico umano. Nel contesto entomologico forense, l'analisi del DNA umano può essere sfruttata anche a livello investigativo e nel cercare di mettere in luce lo stato di negligenza in cui vivono alcuni anziani. In questo lavoro, anche se non siamo stati in grado di ottenere un profilo genetico di altri membri della famiglia dai pidocchi campionati al momento del sopralluogo nella casa di un'anziana Signora morta per cause sconosciute, per la prima volta, la tecnologia *NGS* insieme ad un'accurata ispezione microscopica ed alla strategia dell'incisione dorsale, ci ha consentito di stabilire l'approccio migliore per ottenere un profilo del DNA umano dai pidocchi. L'utilizzo della tecnologia *NGS* in associazione con il *kit HID-Ion AmpliSeq™ Identity Panel (Applied Biosystems®)* ha permesso di ottenere un profilo genetico dal sangue presente nel tratto gastrointestinale degli insetti. Il confronto del profilo genetico estrapolato dagli insetti con quello del campione di riferimento ha confermato una completa sovrapposizione dei due profili, anche quando eventi stocastici sono stati osservati.

Particolare attenzione deve essere prestata quando si verificano questi effetti stocastici. In questi casi, lavorare con materiale di riferimento di buona qualità ed allestire reazioni di sequenziamento in duplicato ed indipendenti potrebbe migliorare l'affidabilità dei dati ottenuti come dimostrato in precedenza (*Salata et al. 2015*). I risultati ottenuti hanno evidenziato la sensibilità della tecnologia *NGS* rispetto all'analisi *STR*

classica anche quando viene applicata a campioni entomologici con DNA in bassa quantità e degradato ed hanno inoltre dimostrato il supporto che tali risultati possono fornire all'analisi entomologica classica.

BIBLIOGRAFIA

- Amendt, J. Krettek, R. & Zehner, R.** 2004. Forensic Entomology. *Naturwissenschaften* , 91: 51-65.
- Boehme, P. Amendt, J. & Zehner, R.** 2012. The use of COI barcodes for molecular identification of forensically important fly species in Germany. *Parasitols Res* , 2325-2332.
- Catts, E. & Goff, M.** 1992. Forensic Entomology in criminal investigation. *Annu. Rev. Entomol* , 37: 253-272.
- Chavez-Briones, M. Hernandez-Cortes, R. Diaz-Torres, P. Niderhauser- Garcia, A. Ancer-Rodriguez, J. Jaramillo-Rangel, G. et al.** 2013. Identification of Human Remains by DNA Analysis of the Gastrointestinal Contents of Fly Larvae. *J Forensic Sci* .
- Clery, J. M.** 2001. Stability of prostate specific antigen (PSA), and subsequent Y-STR typing, of *Lucilia* (*Phaenicia*) *sericata* (Meigen) (Diptera: Calliphoridae) maggots reared from a simulated postmortem sexual assault. *Forensic Science International* 15;120(1-2):72-6
- Di Luise, E. Magni, P. & Sarano, L.** 2007. Gli insetti al servizio degli inquirenti. *Rassegna dell'Arma* .
- DiZinno, J.A. Lord, W.D. Collins-Morton, M.B. Wilson, M.R. & Goff, M.L.** 2002. Mitochondrial DNA sequencing of beetle larvae (Nitidulidae:Omosita) recovered from human bone. *J Forensic Sci*, 47(6) 1337-39.
- Gillariourtua, M. Salona Bordas, M.I. Cainè, L.M. Pinheiro, F. & de Pancorbo, M.M.** 2013. Cytochrome b as a useful tool for the

- identification of blowflies of forensic interest (Diptera, Calliphoridae). *Forensic Science International*, 228(1-3) 132-136.
- Jang, M. Ryu, S. M. Kwon, S.C. Ha, J.-O. Kim, Y.-H. Kim, D.-H. et al.** 2013. A case of oral myiasis caused by *Lucilia sericata* (Diptera: Calliphoridae) in Korea., *Korean J. Parasitol.* 51 (2013) 119–2
- Johnston, E. Ames, C.E. Dagnall, K.E. Foster, J. Daniel, B.E.** 2008. Comparison of presumptive blood test kits including hexagon OBTL., *J. Forensic Sci.* 53 687–9
- Karafet, T.M. Mendez, F.L. Meilerman, M.B. Underhill, P.A. Zegura, S.L. Hammer, M.F.** 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree., *Genome Res.* 18 830–8
- Kidd, K. Kidd, J. Speed, W. Fang, R. Furtado, M.R. et al.** 2012. Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Sci Int Genet* , 6:646-52.
- Li, X. Cai, J. Guo, Y. Xiong, F. Zhang, L. Feng, H. et al.** 2011. Mitochondrial DNA and STR analyses for human DNA from maggots crop contents: A forensic entomology case from central-southern China. *Tropical Biomedicine* , 333-338.
- Linville, J.G. Hayes, J. & Wells, J.** 2004. Mitochondrial DNA and STR Analyses of Maggot Crop Contents: Effect of Specimen Preservation Technique. *J Forensic Sci* 49(2): 341-4.
- Lord, W.D. DiZinno, J.A. Wilson, M.D. Budowle, B. Taplin, D. & Meinking, T.L.** 1998. Isolation, Amplification and Sequencing of

- Human Mitochondrial DNA Obtained from Human Crab Louse, Pthirus Pubis Bloos Meals. *J Forensic Sci* , 43:1097-1100.
- Lord, W. & Stevenson, J.** 1986. *Directory of forensic entomologist, 2nd ed.* Washington DC: Defence pest management Information Analysis centre, Walter Reed army Medical Center.
- Magni, P. Massimelli, M. & Messina, R.** 2008. *Entomologia forense. Gli insetti nelle indagini giudiziarie e medico-legali.* Torino: Edizioni Minerva Medica.
- Mazzanti, M. Alessandrini, F. Tagliabracci, A. Wells, J. & Campobasso, C.** 2010. DNA degradation and genetics analysis of empty puparia: Genetic identification limits in forensic entomology. *Forensic Science International* , Vol 195 : 99-102.
- Mowlavi, G. Nateghpour, M. Teimoori, S. Amin, A. Noohi, F. Kargar, F.** 2011. Fatal nosocomial myiasis caused by *Lucilia sericata*., *J. Hosp. Infect.* 78 2011 338–9
- Mumcuoglu, K.Y. Gallili, N. Reshef, A. Brauner, P. & Grant, H.** 2004. Use of human lice in Forensic Entomology. *J. Med. Entomology* , 803-806.
- Pakstis, A. Speed, W.C. Fang, R. Hyland, F. Furtado, M. Kidd, J. et al.** 2010. SNPs for a universal individual identification panel. *Human Genetics* , 127:325-24.
- Phillips, C. Fang, R. Ballard, D. Fondevila, M. Harrison, C. Hyland, F. et al.** 2007. Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel, *Forensic Sci. Int. Genet.* 1 180–185.
- Replogle, J. Lord, W.D. Budowle, B. Meinking, T.L. & Taplin, D.** 1994. Identification of host DNA by amplified fragment length

polymorphism analysis: preliminary analysis of human carb louse (Anoplura: Pediculidae) excreta. *Journal of Medical Entomology* , 686-90.

Salata, E. Agostino, A. Ciuna, I. Wootton, S. Ripani, L. Berti, A. 2015. Revealing the challenges of low template DNA analysis with the prototype Ion AmpliSeq™ Identity panel v2.3 on the PGM™ Sequencer, *Forensic Sci. Int. Genet.* 22:25-36.

Sánchez, J. Phillips, C. Borsting, C. Balogh, K. Bogus, M. Fondevila, M. et al. 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* , 27:1713-1724.

Spitaleri, S. Romano, C. Di Luise, E. Ginestra, E. & Saravo, L. 2006. Genotyping of human DNA recovered from mosquitoes found on a crime scene. *International Congress Series 1288* , 574-576.

Wells, J. Introna, F.J. Di Vella, G. Campobasso, C. Hayes, J. & Sperling, F. 2001. Human and Insect Mitochondrial DNA Analysis from Maggots. *J Forensic Sci* , 685-687.

Zehner, R. Amendt, J. & Krettek, R. 2004. STR Typing of Human DNA from Fly Larvae Fed on decomposing bodies. *J Forensic Sci* .

Analisi di due denti inviati in busta anonima ad Edda Ciano

INTRODUZIONE

Il 28 aprile 1945 Benito Mussolini e la sua amante Claretta Petacci furono giustiziati dal partigiano comunista Walter Audisio (nome di battaglia "colonnello Valerio") a Giulino di Mezzegra, località in provincia di Como. Il capo del fascismo e della Repubblica Sociale Italiana si trovava in stato di arresto, catturato a Dongo dai partigiani della 52^a Brigata Garibaldi "Luigi Clerici" comandata da Pier Luigi Bellini delle Stelle. Benito Mussolini e Claretta Petacci si trovavano a Dongo perché – secondo la fonte più accreditata – da lì speravano di fuggire, raggiungendo Como e poi la Svizzera. Fino al 18 aprile Mussolini era stato al Palazzo Feltrinelli di Gargnano, un comune vicino a Brescia, da dove aveva guidato la Repubblica Sociale Italiana, uno stato filo-nazista creato nel settembre del 1943 in seguito all'Operazione Quercia: la missione militare nazista che aveva liberato Mussolini, tenuto prigioniero sul Gran Sasso dopo che nel luglio 1943 una seduta segreta del Gran Consiglio del Fascismo ne aveva deciso l'arresto, reso effettivo in seguito ad un ordine del re Vittorio Emanuele III. Mussolini restò a capo della Repubblica Sociale Italiana (RSI) – nota anche come "Repubblica di Salò", perché da Salò, un paese sul lago di Garda, partivano i comunicati del MinCulPop (Ministero della Cultura Popolare), responsabile della propaganda filo-nazista – per quasi due anni: precisamente fino al 22 aprile del 1945 quando, di fronte agli ultimi ufficiali della Guardia Repubblicana – una sorta di polizia interna della

RSI –, pronunciò il suo ultimo discorso pubblico (che secondo alcune fonti terminò con le parole “se la patria è perduta è inutile vivere”). Mussolini restò a Milano fino a quando, in seguito alle insurrezioni del 25 aprile – giorno in cui, da allora, si celebra la Festa della Liberazione – decise di scappare. Nei giorni che trascorse a Milano, Mussolini non era tecnicamente prigioniero ma nemmeno al comando di un esercito. In quei giorni ci furono anche trattative che coinvolsero le forze naziste d’occupazione, Mussolini e alcuni rappresentanti del CLN, cioè il Comitato di Liberazione Nazionale, l’insieme delle forze che dal settembre 1943 lottavano contro il fascismo. Gli storici sembrano concordare sul fatto che in quei giorni a Mussolini fu offerta la possibilità di resa che probabilmente non avrebbe portato a una sua condanna a morte. Mussolini rifiutò soprattutto per non voler passare da traditore, arrendendosi prima dei nazisti.

Rifiutata la resa e scelta la fuga, Mussolini nella sera del 25 aprile – poco dopo l’ordine di insurrezione generale – partì per Como insieme alla sua amante Claretta Petacci, ad alcuni fascisti (accompagnati dalle loro famiglie) e al tenente tedesco Bizier, incaricato da Hitler di scortare Mussolini. Le teorie sul perché Mussolini scelse Como sono diverse: secondo alcuni il suo intento era nascondersi fino all’arrivo degli Alleati, secondo altri l’idea di Mussolini era fin dall’inizio fuggire in Svizzera. Un’altra ipotesi vuole che il piano di Mussolini prevedesse di raggiungere il Ridotto Alpino Repubblicano: il luogo che coincideva più o meno con la Valtellina, da cui gli ultimi fedeli della RSI volevano organizzare la loro ultima difesa. Mussolini, Petacci e gli uomini e le

donne che li accompagnavano passarono da Menaggio e Grandola – due paesi sul lago di Como – e decisero di aggregarsi a un convoglio tedesco formato da circa 200 soldati e alcune decine di autocarri. Nel pomeriggio del 27 aprile Mussolini arrivò a Dongo a bordo di un convoglio di camion. Il convoglio fu fermato durante un controllo da parte dei partigiani: Mussolini fu riconosciuto da alcuni partigiani, disarmato e arrestato dalla 52esima Brigata Garibaldi “Luigi Clerici”. Sulle modalità che portarono alla decisione di uccidere Mussolini, sull’identità di chi lo uccise, così come sull’effettivo luogo della sua fucilazione esistono numerose e tra loro molto diverse versioni. È quasi certo che Mussolini fu ucciso a Giulino di Mezzegra, una frazione dell’attuale comune di Tremezzina, in provincia di Como, a circa 20 chilometri di distanza da Dongo. È anche quasi sicuro che a fucilare Mussolini fu il partigiano Walter Audisio, che procedette alla fucilazione dopo diretti contatti tra la sua Brigata partigiana e più alti esponenti del CLN (Comitato di Liberazione Nazionale) e del CLNAI, il Comitato di Liberazione Nazionale Alta Italia. Ancora più incertezze, dovute alla concitazione di quel momento storico, esistono sul perché – ed eventualmente rispondendo all’ordine di chi – si decise di portare i corpi di Benito Mussolini, Claretta Petacci e di altri diciotto gerarchi della RSI a Milano, in piazzale Loreto, dove vennero appesi a testa in giù quasi nello stesso luogo in cui il 10 agosto 1944 vennero fucilati quindici partigiani antifascisti dalla Legione autonoma mobile Ettore Muti, un corpo militare della RSI.

C’è chi sostiene che il ragioniere Walter Audisio nella notte, in gran segreto, fece scaricare i corpi dei gerarchi fucilati a Giulino di Mezzegra

in un angolo del piazzale. Oltre ai corpi del Duce e della sua amante Claretta Petacci, fucilati personalmente dal «colonnello Valerio», c'erano quello di Marcello Petacci, fratello della donna di Mussolini e definito «ruffiano» dai partigiani e quelli di quindici gerarchi fascisti che stavano con il Duce a vario titolo, chi con l'illusione di salvarsi nella folle fuga verso la Svizzera, chi con la consapevolezza di una testimonianza di fedeltà estrema. L'esposizione dei corpi in piazzale Loreto aveva un valore altamente simbolico e non avvenne per caso, come dichiarò il colonnello Valerio quando, richiamato dal prefetto partigiano Riccardo Lombardi, anche per le proteste del rappresentante americano Charles Poletti e del cardinale Schuster, spiegò che era stato costretto a scaricare i cadaveri dalle tute verdi della Pirelli che non l'avevano lasciato passare. La voce che i corpi di Mussolini, della Petacci e dei gerarchi erano esposti a piazzale Loreto si diffuse presto. Sui cadaveri cominciò lo scempio della folla assetata di vendetta: calci, sputi, anche colpi di pistola. Una donna anziana arrivò a urinare sul volto del Duce. A quel punto venne decisa di issare i corpi sulla tettoia di un distributore di benzina e i corpi vennero appesi a testa in giù. Ci volle la mano pietosa di un prete vicino ai partigiani, don Pollarolo, per chiudere con una spilla la gonna di Claretta e far cessare le urla di scherno. Finalmente quando arrivarono due autocarri, i cadaveri, staccati dal distributore, furono portati all'obitorio dove venne effettuata l'autopsia dal medico di turno il professor Cattabeni. L'autopsia sul cadavere venne eseguita in data 30 aprile 1945. Il professor Cova fece parte dell'equipe di anatomopatologi diretta da Caio Mario Cattabeni che eseguì l'autopsia sul cadavere di

Mussolini. Al termine dell'autopsia, nell'obitorio milanese Cova compilò con scrupolo, ad uso personale, un verbale sull'esame della salma del dittatore fascista che venne reso noto per la prima volta nel settembre del 1994. Nonostante diversi sembrano essere i particolari non riportati nella relazione del dottor Cova, egli precisò che “in bocca mancavano parecchi denti e tutti i superiori di destra” senza però dare ulteriori particolari.

SCOPO DEL LAVORO

Questo lavoro, commissionato dal Ministero per i Beni e le attività culturali, Archivio Centrale dello Stato ed effettuato in collaborazione con l'Istituto Superiore di Sanità ed il Reparto Investigazioni Scientifiche dei Carabinieri di Roma, sezione dei Biologia aveva lo scopo di investigare l'appartenenza a Benito Mussolini di due denti attraverso le analisi utilizzate in ambito forense per l'identificazione umana. Una lettera oltraggiosa contenente questi due denti ed indirizzata ad Edda Ciano, figlia di Benito Mussolini, venne intercettata dagli Americani, consegnata successivamente alle autorità italiane e conservata nell'Archivio Centrale dello Stato fino ai giorni di oggi. Nel 2014 i due denti sono stati consegnati al Reparto Investigazioni Scientifiche di Roma per effettuare le analisi finalizzate ad identificare l'appartenenza dei due denti oggetto di indagine a Benito Mussolini. A tale scopo, i campioni sono stati ispezionati ed è stato effettuato, tramite tampone sterile, un prelievo di eventuale materiale biologico sulla superficie esterna dei denti e successivamente anche la componente biologica interna dei campioni è stata prelevata per le analisi genetiche. Al fine di poter confrontare il profilo genetico ottenuto dai denti con un profilo di confronto noto ottenuto da un familiare di Mussolini, un tampone salivare è stato effettuato ad suo bis nipote.

MATERIALI E METODI

1. Campionamento

I due denti oggetto di indagine sono riportati in Figura 1 .



Figura 1: Molari inviati in lettera anonima alla figlia di Mussolini, Edda Ciano

Al fine di avere un campione di confronto di un familiare ancora in vita di Mussolini, dopo la ricostruzione di un piccolo albero genealogico della

famiglia, un tampone salivare è stato effettuato ad un bis nipote di Mussolini che si è reso disponibile per l'analisi.

2. Ispezione

Un'accurata ispezione dei denti è stata effettuata sia ad occhio nudo al fine di evidenziare eventuali tracce biologiche presenti sulla superficie dei denti stessi sia mediante lo stereomicroscopio (Leica MZ12 –Leica Microsystems srl, Milano) al fine di investigare le caratteristiche morfostrutturali degli denti. Anche un esame radiografico su entrambi i denti è stato effettuato.

3. Test presuntivo e prelievo per analisi genetiche

Dopo un'accurata ispezione, sulla superficie dei denti è stato effettuato il *Combur tritest*. Il *Combur tritest*, comunemente chiamato *Combur-test*, è un *test* presuntivo di tipo colorimetrico utilizzato, in ambito forense, per rilevare la presenza di emoglobina in tracce di presunta natura ematica. Il *test* si basa sulla attività pseudoperossidasi dello ione Fe^{3+} della ferriprotoporfirina-OH che si forma nelle tracce ematiche fuoriuscite dall'organismo. Il saggio consiste nella capacità dello ione Fe^{3+} di catalizzare, in presenza di un idroperossido organico (come il dimetil-diidroperossiesano), l'ossidazione della tetrametilbendizina (TMB) (presente sulla striscia colorimetrica del test) associando questa reazione ad un viraggio cromatico. La TMB funge da riducente mentre l'idroperossido da ossidante. Il *Combur tritest* è un *test* che viene di *routine* utilizzato in ambito clinico per le analisi delle urine. Pur essendo

un *test* molto sensibile, risulta tuttavia poco specifico. La reazione cromatica infatti non è esclusiva dell'emoglobina e può dare luogo a falsi positivi come in presenza di perossidasi vegetale, metalli (in particolare ferro e rame), ruggine, alcuni detergenti contenenti ipocloriti ecc. Inoltre, un ulteriore limite connesso con l'utilizzo di tale tipologia di *test* è il fatto di non essere in grado di discriminare tra sangue umano ed animale: è pertanto un *test* orientativo e non specie-specifico.

Una volta effettuato il *test*, lo strato superficiale dei due denti è stato prelevato mediante tampone sterile ed il materiale biologico raccolto è stato sottoposto ad estrazione.

4. Pulizia e polverizzazione dei campioni

Successivamente al prelievo, la superficie esterna dei denti è stata pulita ed i campioni sono stati sottoposti ad irraggiamento con UV per 45 minuti in *crosslinker* (*Biolink DNA crosslink*). Si è quindi proceduto alla polverizzazione della radice mediante trapano dentistico e circa 50 mg di polvere d'osso per ciascun campione sono stati utilizzati per la successiva fase analitica.

5. Estrazione del DNA

Il protocollo seguito per l'estrazione dalla polvere d'osso è stato pubblicato da *Dabney* e collaboratori (*Dabney et al. 2013*). L'estrazione invece del materiale biologico tamponato sulla superficie dei denti e del tampone salivare di confronto è stata effettuata mediante procedura automatizzata EZ1 seguendo il protocollo *EZ1 DNA Investigator kit*.

6. Quantificazione del DNA

La quantificazione degli estratti, come precedentemente riportato, è stata effettuata mediante *Quantifiler[®] Trio DNA Quantification Kit* (Thermo Fisher Scientific, Oyster Point, CA) che ha permesso di rilevare simultaneamente la quantità totale di DNA umano amplificabile e la quota di DNA maschile eventualmente presente all'interno del campione.

7. Amplificazione e Tipizzazione degli STRs

L'amplificazione del DNA è stata effettuata usando i due *kits* commerciali: *AmpFI STR[®] NGM Select[™] PCR Amplification kit* (Thermo Fisher Scientific) e *AmpFLSTR Yfiler PCR Amplification Kit* (Thermo Fisher Scientific). Questi *kits* di ultima generazione, comprendenti il *set* di marcatori *standard* utilizzati in Europa, amplificano rispettivamente:

- *AmpFI STR NGM Select* (Thermo Fisher Scientific): D10S1248, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D22S1045, D19S433, TH01, FGA, D2S441, D3S1358, D1S1656, D12S391, SE33;
- *AmpFLSTR Yfiler PCR Amplification Kit* (Thermo Fisher Scientific) usato per l'amplificazione degli *STRs* del cromosoma Y comprende i seguenti marcatori: “*European minimal haplotype*” (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393); *Scientific Working Group-DNA Analysis Methods* (SWGDM)-raccomanda il pannello *European minimal haplotype* più DYS438 e DYS439 ed infine ci sono nel *kit* altri loci altamente polimorfici quali

DYS437, DYS448, DYS456, DYS458, DYS635 (Y GATA C4), and Y GATA H4.

Una volta amplificati, i prodotti di PCR sono stati controllati mediante il sequenziatore automatico ABI PRISM 3500 *Genetic Analyzer* (Applied Biosystems) usando il polimero POP4, il capillare di 36 cm e le impostazioni predefinite dello strumento come consigliato dal produttore. I profili genetici sono stati determinati utilizzando il *software data collection 2* e *GeneMapper*[®] ID-X. La soglia analitica è stata fissata a 150 RFU secondo validazione interna.

8. Preparazione libreria *Forenseq*

Una volta quantificati i campioni, è stata allestita una libreria secondo il protocollo *ForenSeq*[™] *DNA Signature Prep* al fine di amplificare i *targets* di interesse a partire dal DNA genomico, in accordo con il manuale d'uso. Il pannello utilizzato (Primer mix B) è composto da 58 *STRs* di cui 27 autosomici, 7 del cromosoma X e 24 del cromosoma Y; 94 *SNPs* identificativi; 56 *ancestry-informative SNPs* e 22 *phenotypic-informative SNPs*.

I campioni sono stati amplificati in duplicato. Gli ampliconi ottenuti sono stati indicizzati e le librerie purificate. Successivamente le librerie sono state normalizzate ed un *pool* 2nM è stato realizzato. Il *pool* è stato quindi fatto correre su *Miseq FGx* Illumina ed i dati analizzati con il *ForenSeq*[™] *Universal Analysis Software*.

RISULTATI E DISCUSSIONE

L'analisi morfostrutturale e radiografica effettuata sui denti ha permesso di evidenziare di essere in presenza di un secondo premolare inferiore di destra ed un terzo molare inferiore di destra. L'elemento dentale numero 1 (premolare) era monoradicolare, ben conservato nella corona e nella radice, con apice intatto. La corona presentava una cuspid vestibolare e due cuspidi linguali con la distale più piccola della mesiale. All'esame ispettivo non si presentavano né carie né ricostruzioni di tipo conservativo. All'esame radiologico si evidenziava una camera pulpare ampia con due ingressi o cornetti pulpari che poi si univano in un unico canale radicolare. L'elemento dentale numero 2 (molare) anch'esso risultava ben conservato nella corona e nelle radici con apici radicolari integri. All'esame ispettivo la corona presentava due cuspidi, due vestibolari e due linguali. Anche in questo caso, non si sono evidenziate carie della corona o ricostruzioni di tipo conservativo. Sulle radici è stato possibile osservare accumuli di tartaro sulla superficie linguale fino alla biforcazione radicolare. Si è pertanto ipotizzato che l'individuo soffrisse di malattia paradontale cronica. All'esame radiografico è stato possibile osservare un'ampia camera pulpare con calcificazioni nella stessa. Il cornetto pulpare mesiale risultava ancora bene evidente mentre quello distale si presentava leggermente retratto. Data la presenza di cuspidi coronali ben conservate e non abrase, camere pulpari ampie con apici intatti, dato il mantenimento di parametri morfometrici in entrambi i denti oggetto di indagine, la presenza di perdita di attacco paradontale e

l'assenza di dentina terziaria di reazione con presenza però di calcificazioni nella camera pulpare dell'elemento dentale numero 2, è stato ipotizzato che i denti potessero appartenere allo stesso soggetto e che l'età dentale potesse corrispondere a quella di una persona adulta tra i 50-60 anni. L'accurata ispezione effettuata sui denti ha permesso di evidenziare la presenza di minuscole tracce di presunta natura ematica sulla superficie esterna dei denti stessi che, dopo aver effettuato il *Combur test* ed aver osservato il viraggio della striscetta colorimetrica da giallo a verde (positività del test), sono state prelevate mediante tampone sterile e sottoposte ad estrazione. Successivamente, previa pulizia della superficie esterna dei denti, questi sono stati sottoposti ad irraggiamento con UV per 45 minuti in *crosslinker* e poi polverizzati con trapano dentistico. 50mg di polvere è stata estratta ed il DNA è stato quantificato. I risultati di quantificazione sono riportati in tabella (Tabella 1).

QUANTIFILER TRIO KIT				
Campioni	Quantità (ng/µl)			ID
	LAT	SAT	Y Target	
D1e	0.040	0.075	0.015	1,8
D2e	0.025	0.050	0.011	2,00
D1	0.060	0.1	0.00	1,6
D2	0.045	0.080	0.00	1,77
ENC	0.00	0.00	0.00	

Tabella 1: Risultati della quantificazione con il *Quantifiler Trio kit*. D1e = prelievo della superficie esterna del dente 1; D2e = prelievo della superficie esterna del dente 2; D1 = dente 1 estrazione da polvere d'osso; D2 = dente 2 estrazione polvere d'osso, ENC = controllo negativo di estrazione e ID = indice di degradazione.

Come si può osservare dalla Tabella 1: nessuna contaminazione si è verificata nel controllo negativo di estrazione e nessuna quantificazione è stata riscontrata per il *target Y* dei campioni D1 e D2 facendo questo ipotizzare di essere in presenza di un individuo di sesso femminile. Diverso risultato invece è stato ottenuto per le tamponature di presunto sangue effettuate all'esterno dei due denti: in entrambi gli estratti è stato infatti possibile osservare la presenza di quantificazione per il *target Y*, facendo questo ipotizzare la presenza di una componente maschile nell'estratto, dato lo sbilanciamento osservato nei valori di quantificazione ottenuti. L'indice di degradazione, i cui dati sono riportati in Tabella 1, è stato calcolato ed i valori ottenuti rientrano in un *range* fra 1,6-2,0. Come riportato nel manuale del *kit* di quantificazione, valori di indice di degradazione compresi fra 1 e 10 indicano la presenza

di una degradazione, seppur moderata, del materiale genetico quantificato. Dopo aver effettuato la quantificazione, gli estratti dei quattro campioni sono stati amplificati mediante *AmpFLSTR® NGM SElect™ PCR Amplification Kit* (Thermo Fisher Scientific, Oyster Point, CA). Dal momento che un profilo biologico misto maschile/femminile, con componente maggioritaria femminile, è stato ottenuto dai campioni D1e e D2e, gli estratti di questi due campioni sono stati amplificati anche utilizzando l'*AmpFLSTR Yfiler PCR Amplification Kit* in grado di amplificare solo la componente minoritaria maschile. Il profilo Y STR ottenuto da entrambi i campioni è risultato completo e perfettamente sovrapponibile. Il confronto effettuato fra i profili Y dei due campioni e quello ottenuto dal campione di confronto del bis nipote di Mussolini ha evidenziato una non compatibilità fra i profili oggetto di indagine. La componente maggioritaria femminile ottenuta da entrambi i campioni è risultata perfettamente sovrapponibile. Dall'analisi dei campioni D1 e D2 è stato estrapolato uno stesso profilo femminile compatibile con quelli (componente maggioritaria) ottenuti dai prelievi effettuati esternamente ai denti. Al fine di ottenere maggiori informazioni circa l'individuo femminile a cui potevano essere appartenuti i denti, è stato deciso di testare il *kit* messo in commercio da Illumina, specifico per il forense, in grado anche di dare informazioni circa i tratti fenotipici (22 SNPs) e l'origine biogeografica (56 SNPs). Partendo dagli estratti è stata allestita una libreria Illumina, purificata, normalizzata e creato un *pool* 2nM per la corsa. L'analisi dei dati, effettuata tramite il *ForenSeq™ Universal Analysis Software*, ha permesso di evidenziare quanto riportato in Figura 2.

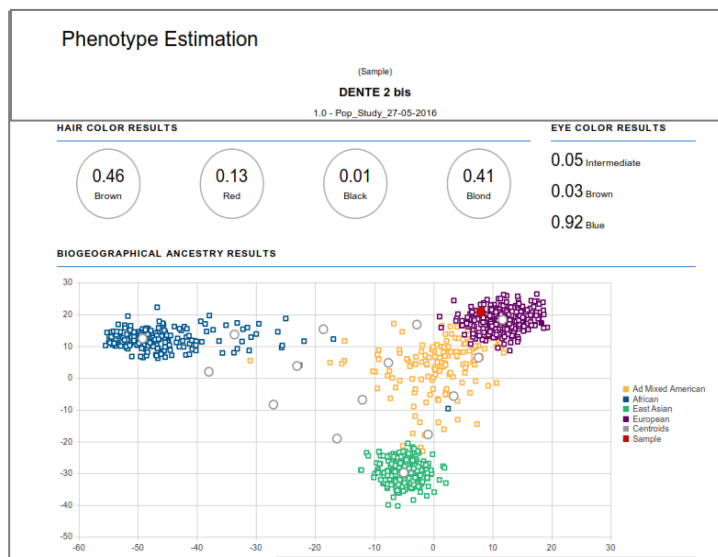


Figura 2: Rappresentazione grafica dei risultati ottenuti dall'analisi con il *ForenSeq™ Universal Analysis Software*.

Come si può evincere dalla Figura 2, i risultati ottenuti in merito al colore dei capelli hanno evidenziato una non chiara attribuzione tra il colore biondo (41%) e quello castano (46%), facendo questo ipotizzare di essere in presenza di un soggetto con capelli castano chiaro. Diverso risultato invece è stato ottenuto per il colore degli occhi per cui è stato possibile effettuare una chiara attribuzione (92%) al colore azzurro degli occhi. Come atteso, l'analisi biogeografica dei campioni analizzati ha indicato l'appartenenza del nostro campione alla popolazione Europea. Considerando il profilo femminile ottenuto e le caratteristiche fenotipiche osservate, non essendo fino ad oggi purtroppo stato possibile avere alcun campione di confronto per l'identificazione, è stato possibile ipotizzare,

dato anche il modo in cui Benito Mussolini e la sua amante Claretta Petacci sono stati uccisi ed i loro corpi esposti insieme agli insulti della folla, che i denti inviati in busta anonima alla figlia di Mussolini, Edda Ciano, possano in realtà essere appartenuti a Claretta Petaccia.

CONCLUSIONI

In questo lavoro, l'analisi forense classica su due denti inviati tramite lettera anonima alla figlia di Mussolini, Edda Ciano, intercettata dagli Americani e conservata nell'Archivio Centrale dello Stato, è stata condotta al fine di valutare l'appartenenza degli stessi a Benito Mussolini. A tale scopo, un tampone salivare è stato prelevato ad un suo bis nipote previo consenso dello stesso, come campione di confronto. In seguito ad un'accurata ispezione, il *combur test* è stato effettuato sulla superficie esterna dei due denti al fine di valutare la presenza di presunte tracce di natura ematica. Data la positività del *test*, entrambi le superfici dei denti sono state tamponate con tampone sterile al fine di prelevare il materiale biologico presente e sottoporlo ad estrazione. Una volta puliti i denti, anche la parte interna degli stessi è stata sottoposta ad estrazione ed a successiva tipizzazione. I risultati ottenuti hanno permesso di evidenziare la presenza di una miscela genetica ottenuta dalla parte esterna dei denti e di un profilo femminile singolo dalla componente interna. L'analisi degli *STRs* del cromosoma Y sui prelievi effettuati esternamente ai denti ha reso possibile l'ottenimento di un profilo Y non sovrapponibile con quello del campione di confronto facendo questo ipotizzare che i profili ottenuti non siano imparentati per linea paterna. Al fine di riuscire ad ottenere maggiori informazioni possibili circa il profilo femminile, data l'assenza di un campione di confronto, una libreria Illumina è stata allestita. I risultati ottenuti hanno permesso di definire che il soggetto femminile aveva gli occhi azzurri ed apparteneva alla

popolazione europea. Minori informazioni è stato possibile ottenere circa il colore dei capelli a causa di una non chiara attribuzione al colore marrone o al colore biondo. I dati ottenuti ed il modo in cui Benito Mussolini e la sua Amante Claretta Petacci sono stati uccisi ed esposti a Piazzale Loreto alle botte, agli insulti ed agli sputi della folla ci fanno ipotizzare che i denti possano essere appartenuti alla giovane amante di Mussolini. Purtroppo però fino ad oggi, in assenza di un campione di confronto, non è stato ancora possibile attribuirli a Claretta Petacci. Nonostante i risultati ottenuti non siano stati utilizzati per un'attribuzione dei denti ad alcun soggetto a causa della mancanza di un campione di confronto idoneo, l'impiego della tecnologia *NGS*, associata all'amplificazione contemporanea di 78 marcatori *SNPs* specifici per investigare i tratti fenotipici e biogeografici, ha permesso di fornire informazioni utili in tutti quei casi in cui non esista un campione di confronto e difficili da ottenere con la tecnologia forense classica a causa dell'elevato numero di *SNPs* investigati contemporaneamente.

Next generation sequencing e DNA capture: il genoma mitocondriale dei calci di Pompei

INTRODUZIONE

1. Cenni storici

La nascita di Pompei risale al VII secolo a.C. in uno scenario caratterizzato da una redistribuzione dell'apparato urbanistico dei centri già esistenti nella Valle del Sarno e da una forte influenza da parte del mondo greco coloniale, nonostante Pompei non sia mai stata sotto il controllo politico e militare del mondo ellenistico. Alcune testimonianze fanno risalire la nascita di questo importante centro urbano già alla fine del IX secolo a.C. quando il popolo degli *Opici* si insediò in forma non stanziale alle pendici del Vesuvio. Secondo la mitologia greca, invece, la città fu fondata in onore di Eracle, eroe greco, che fu omaggiato per le sue gesta dal popolo locale con una sacra processione, definita *pompa*, da cui derivò il nome della città. L'ipotesi più accreditata invece fa risalire l'etimologia del nome della città al termine coniato dal popolo degli *Osci* per indicare il primo insediamento stanziale nella Valle del Sarno risalente al VIII secolo a.C. insediamento che era originariamente costituito da cinque quartieri, chiamati appunto *púmpe*. Rimane tuttavia una delle tante ipotesi in quanto non furono mai rinvenute tracce di tali insediamenti che potessero confermare questa effettiva origine del nome. Secondo una ricostruzione iniziale della città solo a partire dal V secolo a.C. la città cominciò ad ampliarsi, espandendo i propri confini e

delimitandoli con un'imponente cinta muraria. Tuttavia scoperte più recenti hanno mostrato come già a partire dalla metà del VI secolo a.C. la città fosse più estesa di quanto ritenuto in precedenza e fosse già protetta da una prima fortificazione. Il massimo splendore Pompei lo attraversò nel III e nel II secolo a.C. quando la città si trovava sotto l'egemonia Sannita in seguito alla stipula di un trattato di alleanza con il popolo Romano al termine della Seconda Guerra Sannitica. Questa epoca viene definita “il periodo d'oro della città”; Pompei infatti raggiunse la sua massima espansione: venne delineato il perimetro definitivo della città, furono rafforzate le mura di cinta, vennero costruiti edifici pubblici e privati, come teatri, templi dedicati a culti stranieri (come il tempio dedicato alla divinità egizia Iside) ed edifici utilizzati per la formazione fisica, culturale e militare dell'*élite* locale. La zona del *Foro* fu completamente rinnovata e la maggior parte delle *tabernae* (le antiche botteghe commerciali) presenti furono sostituite dalla costruzione di edifici ispirati alle opere monumentali di Roma. In ambito commerciale si sviluppò la produzione di vino e olio e l'esportazione degli stessi si espanse fino in Provenza ed in Spagna.

Nell'89 a.C., in seguito alla Guerra Civile che vide coinvolta direttamente la città, Pompei entrò definitivamente sotto il controllo Romano, divenendo *municipium* prima e *Colonia Cornelia Veneria Pompeianorum* poi. In quest'epoca si assistette ad una vera e propria romanizzazione dell'apparato urbanistico e del profilo architettonico fino a che, intorno al 27 a.C., con la salita al potere di Augusto, Pompei fu considerata una vera e propria “residenza di villeggiatura” del patriziato romano. Fino al 59 d.C. non si registrarono eventi significativi nella storia di Pompei ad

eccezione di un periodo di crisi che subì la città sotto l'impero di Caligola di cui però rimane sconosciuta la causa. La tranquillità dei cittadini di Pompei fu però interrotta il 5 febbraio del 62 d.C. a causa di un violento terremoto che ridusse la città ad un cumulo di macerie. L'episodio fu narrato da Seneca nel sesto libro delle *Questioni Naturali* dedicato all'amico Lucilio. Il terremoto ebbe un forte impatto negativo sulla città risentendone, in primo luogo, il commercio e l'economia che calarono bruscamente. Molti personaggi illustri inoltre si trasferirono in altre località temendo per la propria incolumità. Negli anni successivi l'attenzione fu rivolta all'attività di ricostruzione della città per la messa in sicurezza di edifici pubblici e delle dimore private del patriziato romano. Tuttavia le opere di ricostruzione di edifici pubblici e privati non erano ancora state ultimate quando Pompei fu sepolta definitivamente dall'eruzione vulcanica del 79 d.C.

1.1 L'eruzione del 79 d.C.

L'evento che segnò la fine di Pompei fu l'eruzione vulcanica del 79 d.C., considerata come la più famosa eruzione vulcanica della storia che causò migliaia di vittime. Il 24 agosto del 79 d.C., dopo circa otto secoli di inattività, il vulcano Somma, erroneamente quando si parla dell'eruzione del 79 d.C. si attribuisce questo evento al Vesuvio, si risvegliò.

La Figura 1 mostra l'attuale cratere del Vesuvio che, come si può osservare, è circoscritto da una mezzaluna che, per chi guarda una classica "cartolina da Napoli", appare come una mezzaluna alla sinistra del Vesuvio. In realtà quella mezzaluna alla sinistra del Vesuvio è ciò che

rimane del cratere dell'antico vulcano Somma. Il Vesuvio pertanto si generò proprio in seguito all'eruzione del 79 d.C. e si innalzò esattamente al centro del cratere del monte Somma impiegando secoli per raggiungere le dimensioni attuali. Il Vesuvio può pertanto essere considerato il “figlio della tragedia di Pompei” e non il suo artefice.



Figura 1: Cratere del Vesuvio visto dall'alto. La mezzaluna che circonda il cratere è ciò che oggi rimane dell'antico monte Somma.

All'alba del 24 Agosto venne descritta una grande nuvola a forma di pino che emerse dalla sommità del vulcano, seguita da una violenta esplosione che fece fuoriuscire dal vulcano le antiche masse laviche sotto forma di lapilli e pomice bianche che cominciarono a depositarsi su tutta la città di Pompei. Gli abitanti iniziarono ad abbandonare le proprie abitazioni e a rifugiarsi in luoghi coperti per evitare di essere colpiti dalle “bombe vulcaniche”, “gocce” di roccia ardente che si formano durante

un'eruzione vulcanica e dal crollo degli edifici causati dal peso degli strati di lapilli che si andavano via via depositando per la continua attività sismica. L'immensa nube che raggiunse l'altezza di 14 km oscurò completamente il paesaggio campano e gli abitanti cominciarono sin da subito a soffrire di crisi respiratorie a causa dell'aria densa divenuta irrespirabile. Intorno alle ore 20.00 del 24 agosto si ebbe l'emissione della colonna eruttiva (cioè una emissione in verticale di frammenti di magma, ceneri e gas che vengono espulsi a causa delle pressioni che si creano all'interno della camera magmatica e che generano una nube che può raggiungere diversi km di altezza nell'atmosfera fino alla stratosfera) che generò i primi due flussi piroclastici che raggiunsero le abitazioni prossime alle pendici del vulcano. Il giorno seguente, nelle prime ore del mattino, Pompei venne colpita da altri flussi piroclastici che, secondo testimonianze storiche, raggiunsero la temperatura di 400 gradi e una velocità massima di 30 m al secondo, distruggendo anche i centri di Ercolano, Boscoreale e Terzigno. Mentre il livello delle pomici saliva fino a 2,50- 2,80 metri, la fuga per mare era impedita da violenti maremoti e dal ritiro delle acque per centinaia di metri. Altri flussi piroclastici si abbattono su Pompei scavalcando le mura di cinta e causando la morte di quanti erano rimasti ancora in città. L'ultima eruzione, intorno alle 08.00 del 25 agosto del 79 d.C. (o 25 ottobre secondo la più recente datazione), determinò definitivamente la fine della civiltà di Pompei: la colata lavica seppellì completamente gli edifici che ancora emergevano dallo strato di lapilli che aveva già raggiunto i 3 metri di altezza. L'attività eruttiva cominciò poi ad affievolirsi ma la pioggia di

ceneri proseguì per i quattro giorni successivi, fino a che l'attività del Vesuvio non cessò del tutto.

1.2 Gli scavi archeologici

I primi tentativi di riportare alla luce l'antica città di Pompei si ebbero già qualche anno dopo l'eruzione sotto l'imperatore romano Alessandro Severo (intorno al 222 d.C.). Il tentativo tuttavia fallì a causa dello spesso strato di cenere e lapilli e della rigogliosa vegetazione che nel frattempo aveva ricoperto tutta l'area. È solo intorno agli anni 1594-1600 che le attività per riportare alla luce gli scavi ripresero. La scoperta dei resti della città fu del tutto casuale e risale a quando l'architetto Domenico Fontana fu incaricato dal conte di Sarno, Muzio Tuttavilla, di costruire un canale per portare le acque del fiume Sarno a Torre Annunziata. Durante i lavori di costruzione del canale furono rinvenuti i resti di edifici, iscrizioni e monete che però non furono immediatamente attribuite alla città di Pompei. Il terremoto poi del 1631 impedì il proseguo dei lavori.

I veri e propri scavi per riportare alla luce la città di Pompei iniziarono nel 1738 per volere di Carlo di Borbone, Re delle Due Sicilie, che affidò la direzione dei lavori all'ingegnere Alcubierre. Gli scavi furono condotti in maniera non sistematica in quanto il principale intento era quello di riportare alla luce solo gli oggetti preziosi utili per arricchire il museo di Portici. Intorno agli anni 1759-1799 sotto Ferdinando I fu riportata alla luce parte della città adottando un sistema di scavo tale da permettere di lasciare gli affreschi parietali sui muri originari senza la necessità di asportarli e di trasferirli altrove. Successivamente, negli anni seguenti, furono riportati alla luce parte della zona dei teatri, del tempio di Iside e

del Foro Triangolare; tra il 1760 e il 1772 l'attenzione fu rivolta alla zona nord-occidentale della città dove furono rinvenuti la Villa di Diomede, la Casa del Chirurgo e Via dei Sepolcri e dove furono scoperti diciotto corpi, morti a causa dell'eruzione. Intorno ai primi anni del 1800 Giuseppe Bonaparte e Gioacchino Murat incentivarono le opere di scavo, individuando la cinta muraria della città e portando alla luce la strada tra la Villa di Diomede e la Casa di Sallustio, il Foro e la Basilica. È in questi anni che la Pompei antica divenne un vero e proprio “museo all'aperto”. Con l'unità di Italia (1861), i lavori furono affidati a Giuseppe Fiorelli che impostò, con un approccio sistematico, l'organizzazione dei lavori, suddividendo la città in *regiones*, cioè quartieri, ed in *insulae*, ovvero agglomerati di case (Figura 2).

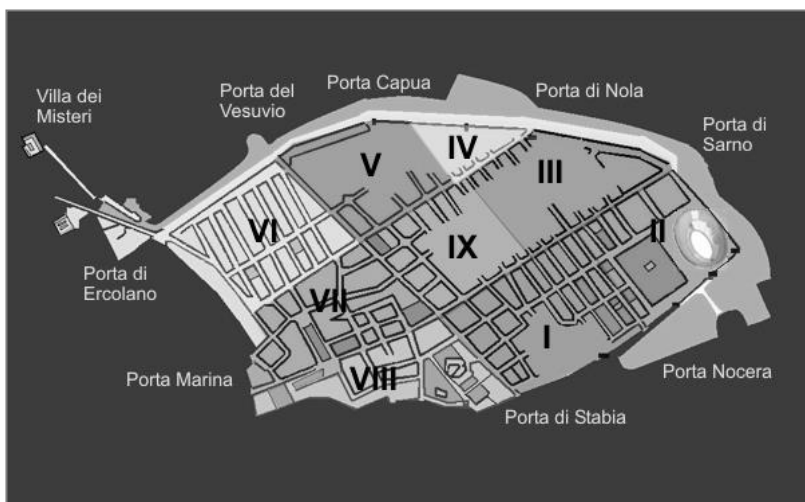


Figura 2: Suddivisione della città di Pompei in *regiones* secondo l'organizzazione proposta da G. Fiorelli.

L'archeologo Fiorelli, oltre ad aver progettato questo sistema di suddivisione e catalogazione della città, fu il primo ad introdurre la tecnica del calco in gesso. Questa tecnica si basava sul far colare una miscela di gesso liquido e acqua all'interno dei vuoti lasciati, negli strati di cenere indurita, dalla decomposizione dei corpi. Una volta solidificato, il calco in gesso riproponeva esattamente il volume, la forma e la posizione dell'oggetto o del corpo che era stato sepolto. È sorprendente come, con questa tecnica, sia stato possibile riuscire ad imprimere nel calco le fattezze, l'aspetto esteriore e i gesti di chi aveva vissuto quei tragici momenti. Inoltre il gesso indurito può aver incluso anche parti di ossa, tessuti e altro materiale permettendone la conservazione nel corso degli anni (Figura 3).



Figura 3: Rappresentazione di alcuni dei calchi in gesso riprodotti con la tecnica messa a punto dall'archeologo Fiorelli.

Tra il 1924 e il 1961 i lavori furono diretti da Amedeo Maiuri il quale riportò alla luce edifici del calibro della Villa dei Misteri e, sempre in quel periodo, fu delineato il perimetro della città e si iniziò a scavare

negli strati sottostanti al livello del 79 d.C. con l'intento di riscoprire la storia più antica di Pompei. Il terremoto dell'Irpinia del 1980 rallentò però i lavori di scavo che ripresero solo in seguito, grazie ai fondi stanziati dal Fondo Investimenti e Occupazione per la valorizzazione dell'area archeologica. Negli anni compresi fra il 1987 e il 1989 nell'ottica della promozione del Progetto Neapolis -progetto sistema per la valorizzazione integrale delle risorse ambientali ed artistiche dell'area vesuviana- furono catalogati beni archeologici come pitture e mosaici e censiti i beni artistici, ambientali e territoriali dell'intera area vesuviana ricorrendo anche all'ausilio di avanzate tecnologie informatiche. Nel 1997 l'area archeologica di Pompei, Ercolano e Torre Annunziata venne dichiarata dall'UNESCO patrimonio dell'umanità. A partire dal 2012 è stato promosso il progetto “Grande Progetto Pompei” grazie all'accordo stipulato tra il Ministro della Coesione, il Ministro dei Beni e delle Attività Culturali, il Ministro dell'interno, il Ministro dell'Istruzione, dell'Università e della Ricerca ed il Presidente dell'Autorità di Vigilanza sui contratti pubblici per la sicurezza degli appalti. I punti cardine di questo progetto si basavano sulla riqualificazione del sito archeologico di Pompei ed in particolare si basavano sulla riduzione del rischio idrogeologico con la messa in sicurezza dei terrapieni non scavati, sulla messa in sicurezza delle *insulae*, sul consolidamento e restauro delle murature, sul consolidamento e restauro delle superfici decorate, sulla protezione degli edifici dalle intemperie con conseguente aumento delle aree visitabili e sul potenziamento del sistema di videosorveglianza.

Gli scavi archeologici effettuati negli anni hanno riportato alla luce circa 1150 corpi rinvenuti negli strati di lapilli inferiori, probabilmente rimasti vittime del crollo degli edifici per il peso dei materiali vulcanici che si erano via via stratificati sui tetti; mentre altri 650 corpi furono rinvenuti nella parte superiore dei depositi piroclastici morti probabilmente a causa delle nubi ardenti della seconda fase dell'eruzione.

SCOPO DEL LAVORO

In questo lavoro il *Next Generation Sequencing* ed il metodo del *target enrichment capture* sono stati applicati allo studio di campioni ossei altamente degradati prelevati, al momento del restauro, dai calchi in gesso rinvenuti nel sito archeologico di Pompei in due differenti abitazioni al fine di indagare le eventuali relazioni di parentela esistenti fra gli individui campionati. Nella prima abitazione sono stati rinvenuti due individui adulti abbracciati che gli archeologi hanno supposto potessero essere imparenti per linea materna: due giovani sorelle (C-33 e C-32). Nella seconda abitazione sono stati invece rinvenuti quattro differenti individui che gli archeologi hanno ipotizzato potessero appartenere ad un unico gruppo familiare ben definito: un individuo adulto di sesso maschile, il presunto padre (C-21), una donna adulta, la presunta madre (C-34) e due infanti (C-23 e C-35). I campioni analizzati in questo lavoro di tesi si presentavano in condizioni non ottimali per le analisi genetiche sia in termini qualitativi, le ossa erano per lo più fuse con la calce, che in termini quantitativi, i campioni a disposizione erano di dimensioni estremamente ridotte. Il lavoro è consistito inizialmente nella rimozione della calce, nella pulizia e polverizzazione degli elementi scheletrici e nella estrazione del DNA in essi contenuto. Successivamente alla costruzione della libreria, il genoma mitocondriale è stato catturato mediante l'utilizzo di specifiche sonde e la sequenza consenso ricostruita mediante mappaggio di tutte le *reads* ottenute sulla sequenza di riferimento rCRS al fine di indagare le presunte relazioni di parentela

esistenti fra i due differenti gruppi di individui e verificare, in questo modo, l'appartenenza dei campioni oggetto di indagine alla medesima linea di discendenza materna.

MATERIALI E METODI

1. Campionamento

In questo lavoro sono stati analizzati campioni prelevati ad alcuni dei calchi rinvenuti negli scavi di Pompei al momento del restauro degli stessi. Questo studio ha riguardato in particolare l'analisi di cinque frammenti ossei ed un dente prelevati dai calchi rinvenuti in due edifici riportati alla luce nel corso degli scavi archeologici. Il primo sito in cui sono stati rinvenuti alcuni dei campioni oggetto di studio è la *Casa del Criptoportico* residenza signorile che si apriva in via dell'Abbondanza nell'*insula* 16. Disposta su due livelli, era una *domus* lussuosa dotata di impianto termale e con un grande corridoio sotterraneo voltato, il criptoportico da cui la casa prendeva il nome. Dal criptoportico si poteva accedere alle terme private dell'abitazione e ad un soggiorno (*oecus*). Il criptoportico era decorato con episodi dell'*Iliade* e il suo soffitto con volta a botte era decorato con stucchi. L'*oecus* (soggiorno) presentava un importante pavimento a mosaico ed era decorato con quadretti che riprendevano motivi dionisiaci. Il piano superiore ospitava il triclinio estivo in muratura ricoperta di intonaco di colore rosso. Nell'edificio trovarono la morte dieci individui che, in un momento di stasi dell'eruzione uscirono nel giardino, ormai ricoperto da uno spesso strato di lapilli. Solo per quattro di essi fu possibile eseguire il calco. Fra questi i calchi C-33 (Figura 4) e C-32 (Figura 5), appartenenti, come ipotizzato dagli archeologi, a due giovani donne, presunte sorelle, da cui sono stati

ricavati i campioni (rispettivamente un dente e un frammento osseo) analizzati in questo studio.



Figura 4: Immagini relative al campione C-33. A sinistra è mostrato il calco al momento del restauro rinvenuto nell'abitazione, mentre a destra il dente dal quale sono state fatte le analisi.

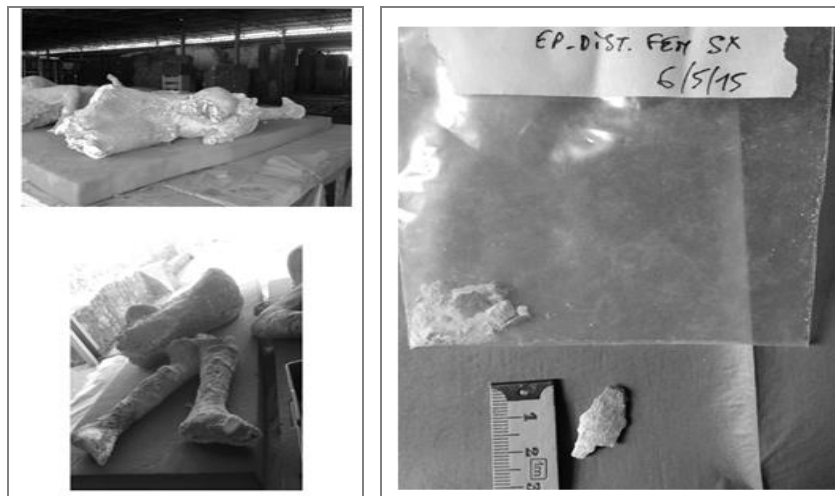


Figura 5: Calco a sinistra e frammento osseo a destra del campione C-32

Il secondo sito da cui sono stati prelevati gli altri campioni è la Casa del Bracciale d'Oro situata nelle immediate vicinanze della Casa di *M. Fabius Rufus* (entrambe avevano l'accesso dal Vico del Farmacista) e localizzata nella regio VI, *insula occidentalis* 17 numero 42. Il nome attribuito a questo edificio derivava dal ritrovamento di un bracciale d'oro finemente decorato rinvenuto ancora al polso di una presunta donna rimasta vittima dell'eruzione. La Casa del Bracciale d'Oro rappresenta un'innovazione architettonica del tempo: è costruita sulle antiche mura della città ed ha un motivo architettonico che fonde le caratteristiche del classico modello romano con quelle della villa suburbana avendo una disposizione in verticale che si articola su tre livelli con ambienti che si affacciano su ampie terrazze panoramiche. Un primo tentativo di riportare alla luce la Villa fu effettuato in età borbonica tra il 1758 e il 1763 anche se gli scavi effettivi cominciarono intorno al 1974 quando si iniziarono a definire i confini dei vari complessi abitativi di quella zona della Pompei antica che fino ad allora venivano genericamente identificati con il nome di "Case di Fabio Rufo". All'interno di questa Villa persero la vita diversi individui, alcuni travolti dal crollo del pianerottolo mentre scendevano le scale che li avrebbero condotti all'esterno, altri, rinvenuti molto vicini fra loro ed ai precedenti, nel piano inferiore della casa. Una delle vittime, "certamente" una donna, certezza attribuita dagli studiosi/archeologi per il bracciale d'oro che aveva al braccio destro, sosteneva con il braccio sinistro un bambino in tenera età; a fianco c'era un adulto, forse un loro congiunto. Oltre al bambino la donna recava, in un sacchetto o cofanetto, un cospicuo

gruzzolo di quaranta aurei e 176 denari d'argento, due anelli d'oro ed una gemma. A pochi metri di distanza si rinvenne un bambino di circa 4 anni di età di cui fu eseguito il calco, insieme ai due individui adulti ed al bambino appena descritti. Il calco del bambino di circa 4 anni fu identificato come C-23 (Figura 6), del presunto maschio adulto come C-21 (Figura 7), della presunta donna adulta come C-34 e dell'altro bambino di tenera età come C-35 (Figura 8).



Figura 6: Immagine relativa al campione C-23; a sinistra è riportato il calco, mentre a destra il frammento osseo utilizzato per le analisi.



Figura 7: Calco e frammento osseo relativi al campione C-21

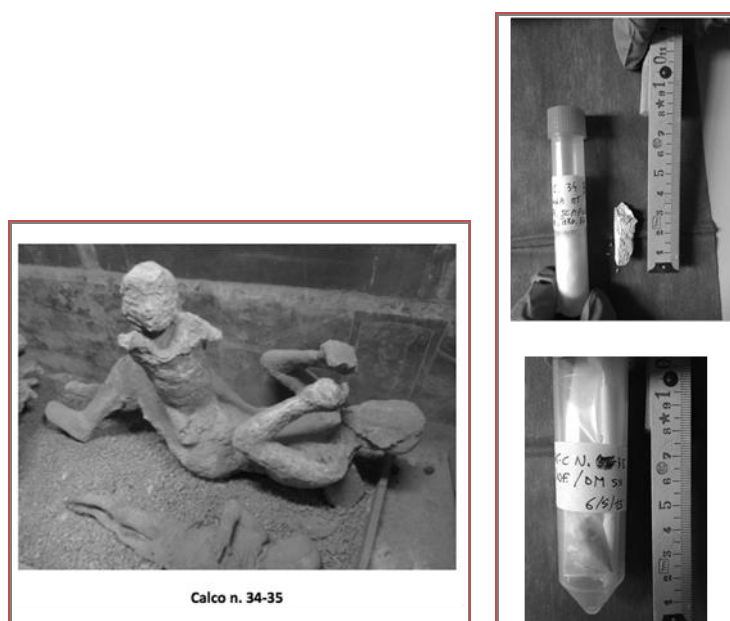


Figura 8: Immagini relative dei calchi e dei frammenti ossei relativi ai campioni C-34 e C-35.

2. Ambiente di lavoro e prevenzione delle contaminazioni

Il lavoro sperimentale è stato condotto presso i Laboratori di Antropologia Molecolare dell'Università degli Studi di Firenze. Al fine di prevenire il rischio di contaminazioni, il laboratorio è dotato di due aree fisicamente separate, con accesso indipendente per seguire le fasi di pre- e post- amplificazione richieste dal protocollo *NGS*. Nell'area di PRE-amplificazione sono state eseguite le operazioni di pulizia e polverizzazione dei campioni, estrazione del DNA e preparazione delle librerie. In quest'area le superfici, gli strumenti, i materiali e la vetreria sono stati pre-trattati con ipoclorito di sodio e sottoposti a irradiazione con raggi UV a 254 nm per almeno 45 minuti prima e dopo l'utilizzo. Particolare attenzione viene rivolta anche ai reagenti che vengono di *routine* suddivisi in aliquote al fine di evitarne l'uso prolungato su un numero elevato di campioni. Controlli negativi sono stati aggiunti a tutti i passaggi analitici. Nell'area POST sono state eseguite le fasi di quantificazione, arricchimento e cattura del DNA mitocondriale previste dal protocollo *NGS* Illumina.

3. Pulizia e polverizzazione dei campioni

Una volta giunti in laboratorio, i campioni sono stati catalogati, fotografati e accuratamente puliti mediante rimozione di uno strato superficiale di 1-2 mm di spessore tramite trapano dentistico allo scopo di eliminare sia la calce adesa ai frammenti ossei oggetto di indagine che eventuali residui di materiale biologico estranei al campione. Prima dell'utilizzo, le punte metalliche utilizzate per la pulizia dei campioni sono state lavate con acido acetico al 30%, sciacquate in acqua

bidistillata e sterilizzate sotto UV per almeno 45 minuti. Una volta rimosso lo strato superficiale da ciascun frammento osseo, ogni superficie del campione è stata sottoposta a irraggiamento con UV per 45 minuti all'interno di un *crosslinker* (*Biolink DNA crosslink*).

Successivamente i campioni sono stati polverizzati mediante l'utilizzo di un trapano dentistico e 50 mg di polvere d'osso di ciascun campione sono stati prodotti per la successiva fase analitica.

4. Estrazione del DNA da polvere d'osso

La fase di estrazione del DNA si basa sulla progressiva solubilizzazione delle strutture cellulari che proteggono il DNA mediante trattamento della polvere con soluzioni contenenti tensioattivi, agenti chelanti ed enzimi quali le proteasi che sono responsabili rispettivamente della distruzione delle membrane biologiche grazie alla solubilizzazione dei lipidi di membrana, dell'eliminazione dei sali di calcio che costituiscono la struttura ossea dei campioni e della denaturazione delle proteine cellulari. Il DNA è stato estratto seguendo il protocollo di *Dabney* e collaboratori (*Dabney et al. 2013*). Il protocollo, specifico per i campioni altamente degradati, ha permesso il recupero anche di frammenti corti di DNA (presunto materiale endogeno).

5. Quantificazione del DNA estratto

Una volta ottenuto l'estratto, quest'ultimo è stato quantificato, come precedentemente riportato, mediante il *Quantifiler[®] Trio DNA Quantification Kit*. Questo tipo di analisi ha permesso di quantificare

simultaneamente il DNA nucleare totale umano e la quota di DNA maschile eventualmente presente all'interno del campione.

6. Preparazione delle librerie Illumina

Per la preparazione delle librerie Illumina è stato seguito il protocollo pubblicato da *Meyer e Kircher (Meyer and Kircher, 2010)*. Il protocollo si articola in tre fasi: 1. la fase di *Blunt-end Repair* per riparare i danni all'estremità dei frammenti, 2. la fase di *Adapter Ligation* che consente il legame degli adattatori alle estremità dei frammenti ed infine 3. la fase di *Adapter Fill-in* per riparare eventuali *nicks* presenti fra adattatori e inserto di DNA (Figura 9). Successivamente è stata effettuata la fase di *Indexing PCR* che ha permesso di indicizzare e al contempo arricchire le librerie.

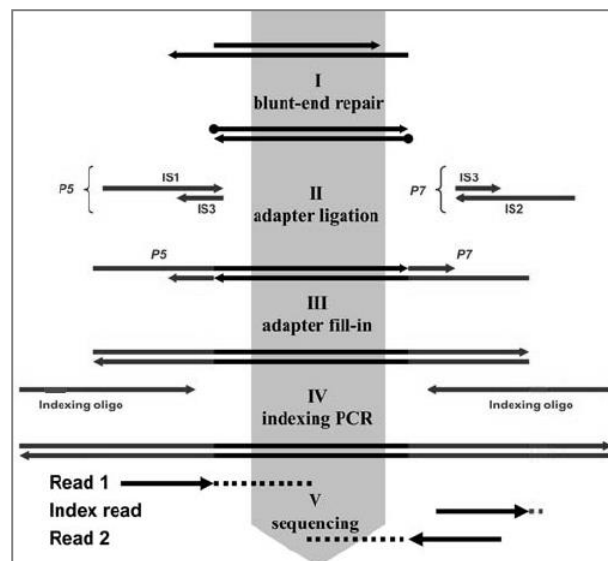


Figura 9: Fasi della preparazione delle librerie Illumina secondo il protocollo proposto da *Meyer e Kircher (2010)*

6.1 Blunt-end repair

Questa prima fase di preparazione delle librerie consente la riparazione delle estremità sporgenti al 5' e al 3' dei frammenti di DNA e permette di generare estremità piatte, funzionali per il successivo legame degli adattatori. La reazione è catalizzata dalla T4 polinucleotide chinasi (PNK) e dalla T4 polimerasi: il primo enzima catalizza la formazione di un gruppo fosfato all'estremità 5' del frammento, mentre il secondo guida l'estensione delle estremità a singolo filamento fino a generare estremità piatte (*Blunt-end*). Il prodotto ottenuto è stato successivamente purificato mediante il *MinElute Purification Kit* (Qiagen).

6.2 Adapter ligation

In questa fase si legano all'estremità della molecola di DNA gli adattatori Solexa Illumina specifici per il sequenziamento (P5 e P7). Il protocollo di preparazione della *Solexa Adapter Mix* prevede la formazione di due adattatori (P5 e P7) solo parzialmente a doppio filamento, essendo l'oligonucleotide *IS3_adapter_P5+P7* complementare solo a una porzione dei due oligonucleotidi P5 e P7 che danno il nome agli adattatori. Questa particolare conformazione garantisce che, durante la successiva fase di *Adapter Ligation*, gli adattatori si leghino nel corretto orientamento alle molecole di DNA del campione possedendo un'unica estremità piatta. Una volta avvenuta la reazione, si è proceduto alla purificazione utilizzando il *MinElute Purification Kit* ed eluendo in 20 µl di TET.

6.3 Adapter fill-in

L'*adapter fill-in* ha permesso di eliminare gli eventuali *nicks* che possono essersi formati durante la fase di *Adapter Ligation* tra l'adattatore ed il frammento di DNA. In seguito all'*adapter fill-in* i campioni sono stati sottoposti ad indicizzazione.

6.4 Indexing PCR e qPCR

Nella fase di *indexing* PCR le librerie sono state indicizzate scegliendo combinazioni di indici diverse per ciascun campione; questo ha consentito di sequenziare librerie appartenenti a diversi campioni in un unico *pool* (strategia *multiplexing*) e successivamente discriminare un campione dall'altro durante l'analisi delle sequenze sulla base della coppia di indici assegnata. La scelta della combinazione di indici è un passaggio estremamente importante e delicato per evitare di generare errori nell'acquisizione delle immagini nella piattaforma Illumina durante il sequenziamento e per permettere il riconoscimento esatto dei campioni sequenziati in *pool*. 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 i successivi passaggi di amplificazione e quantificazione. Al termine della reazione le librerie sono state purificate con *MinElute Purification Kit (Qiagen)*, come descritto in precedenza. Successivamente l'efficienza della *Indexing* PCR è stata valutata mediante *Real-Time* PCR (RT-PCR). La *RT-PCR* è una PCR quantitativa (qPCR) che, grazie all'utilizzo di un intercalante, ha permesso di monitorare la fase esponenziale della reazione di PCR e quantificare il

contenuto di DNA presente in ogni libreria. La *mix* di reazione è stata preparata utilizzando il *Kit SsoFast EvaGreen (Bio-Rad)* che utilizza il *SYBR green* come fluoroforo ed i *primers Solexa fwd IS5* e *rvs IS6*, complementari alla sequenza degli *indexing-primers* esterna a quella degli indici.

Alla fine della reazione è stata ottenuta, per ciascun campione e *standard*, una curva di amplificazione, cioè un grafico in cui, sull'asse delle X, viene riportato il numero di cicli della PCR, mentre sull'asse delle Y è riportata l'intensità di fluorescenza che sarà proporzionale alla quantità di amplificato prodotto dalla reazione (Figura 10).

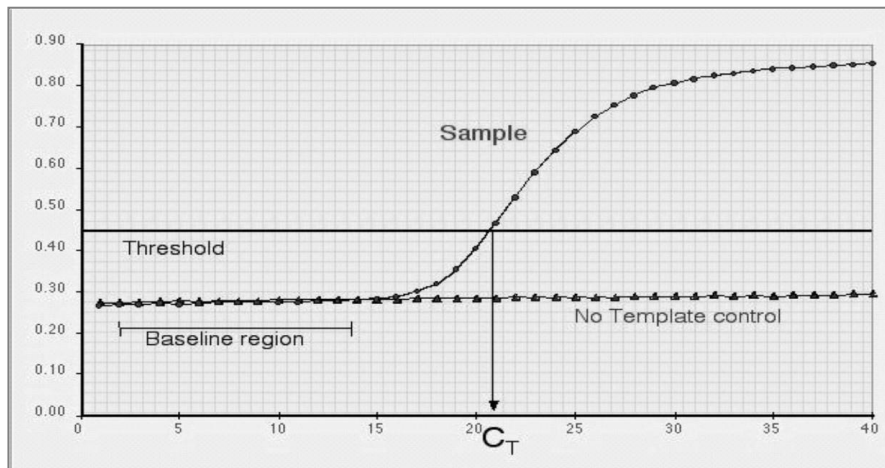


Figura 10: Grafico della reazione di amplificazione. Il plot mostra la variazione dell'intensità di fluorescenza durante i vari cicli di reazione.

Nel grafico riportato si possono osservare quattro distinte fasi del ciclo di amplificazione: una prima fase (*baseline region*) che corrisponde ai cicli

iniziali in cui non si misura una variazione del segnale di fluorescenza; una seconda fase esponenziale, in cui i prodotti di reazione di PCR aumentano di circa il doppio per ogni ciclo e in cui si osserva un conseguente aumento di fluorescenza, una fase lineare, in cui si ha un leggero appiattimento della curva e una fase di plateau in cui l'appiattimento della curva diviene più marcato e rappresenta il punto a partire dal quale la reazione non può più procedere a causa del consumo di tutti i reagenti. Durante la fase esponenziale il *Threshold cycle* (C_T) rappresenta un valore che esprime il numero di cicli oltre il quale si registra un aumento della fluorescenza. Il valore di tale C_T è fortemente dipendente dalla quantità di campione presente all'inizio della reazione di amplificazione. Se all'inizio della reazione è presente una alta concentrazione di campione occorrono pochi cicli per dare un segnale di fluorescenza al di sopra della linea soglia, *Threshold line*, ottenendo così un basso valore di C_T . Al contrario, se all'inizio della reazione la concentrazione del campione è bassa, occorrono più cicli di amplificazione per ottenere un adeguato segnale di fluorescenza e, di conseguenza, il valore C_T aumenta. Il *Threshold cycle* è legato alla concentrazione di DNA iniziale dall'equazione:

$$C_T = \log ([DNA] \text{ iniziale})$$

Riportando i valori di C_T relativi agli *standard* che hanno concentrazioni che vanno dalla 10^8 a 10^3 in funzione del $\log ([DNA] \text{ iniziale})$, si ottiene una retta, detta curva di taratura, che permette di estrapolare la concentrazione dei campioni in esame.

Per avere una quantificazione ottimale e precisa, i seguenti parametri della curva standard devono essere valutati:

- R^2 il coefficiente di correlazione che riflette la linearità della distribuzione dei dati e ci fornisce informazioni anche sulla qualità del risultato; R^2 deve essere maggiore di 0.980.
- Efficienza di amplificazione cioè un parametro legato al coefficiente angolare della retta (la pendenza o *slope*) dalla seguente equazione:

$$E = [10 (-1/slope)] - 1$$

Il valore ottimale dello *slope*, che ci consente di avere un'efficienza del 100%, è pari a -3,32 (*Paternò et al.* 2006).

6.5 Arricchimento della libreria tramite PCR

L'arricchimento delle librerie di DNA è stato effettuato mediante PCR utilizzando il kit *Herculase* (*Agilent Technologies*). I prodotti della reazione sono stati purificati con il *MinElute Purification Kit* (*Qiagen*) ed al fine di stimare quantitativamente e qualitativamente la libreria i campioni sono stati corsi *sull'Agilent Bioanalyzer 2100*, usando il kit DNA1000.

7. Cattura del DNA mitocondriale

Mediante l'utilizzo del protocollo *Maricic* (Maricic *et al.* 2010), il DNA mitocondriale di interesse all'interno delle librerie prodotte è stato selezionato/catturato mediante l'utilizzo di sonde specifiche *home made*.

7.1 Preparazione delle sonde

Le sonde sono state preparate in laboratorio a partire da un prelievo effettuato mediante tampone salivare ad un donatore previo suo consenso informato. L'estrazione del DNA è stata eseguita mediante l'utilizzo del kit *QIAamp Investigator Kit* (Qiagen) secondo il manuale di uso. L'estratto ottenuto è stato quantificato al *NanoDrop 1000* (*Thermo Scientific*) e utilizzato per allestire due *long-range* PCR con il *Kit Expand Long Range System* (Roche), in modo tale da ottenere due amplificati che, sovrapponendosi parzialmente, possano coprire l'intero genoma mitocondriale. I prodotti di PCR sono stati controllati su gel d'agarosio, purificati con il kit *MinElute* e successivamente frammentati mediante sonicazione con il sistema *Bioruptor* per ottenere frammenti di DNA di lunghezza pari a circa 300 bp. I frammenti ottenuti sono stati sottoposti a *blunt-end repair*, purificati mediante *MinElute Purification Kit* (Qiagen) e biotinilati, ovvero legati a specifici adattatori biotinilati (Adaper Bio-T/B), in modo da potersi legare, una volta ibridate con il DNA, alle biglie magnetiche durante la fase di cattura. Dopo purificazione, le sonde sono state misurate al *NanoDrop* al fine di garantire la massima efficienza di interazione tra sonde e biglie magnetiche.

7.2 Interazione delle sonde con le biglie magnetiche

Il legame delle sonde di mtDNA alle biglie magnetiche avviene grazie all'interazione specifica tra biotina e streptavidina presente quest'ultima sulle biglie magnetiche. Le biglie rivestite con le sonde vengono preparate il giorno stesso in cui viene allestito l'esperimento di cattura in quanto la streptavidina associata alle biglie magnetiche è in grado di staccarsi più facilmente quando è legata a molecole di DNA. Un volume di biglie pari a 5 μ l in grado di legare 100-125 ng di sonde sono state utilizzate per la reazione. Dal momento che il protocollo *Maricic* riporta che l'efficienza del legame tra le sonde e gli adattatori e tra questi e le biglie è al massimo del 30%, è necessario disporre di 400-500 ng di sonde biotilate affinché i 5 μ l di biglie siano completamente saturati con le sonde.

7.3 Ibridazione fra la libreria e le sonde

Prima di procedere con la fase di ibridazione fra le sonde e le librerie, quest'ultime devono subire un trattamento preliminare affinché si possano ottenere frammenti di DNA a singolo filamento che non si riappaino fra di loro. Per garantire ciò il protocollo *Maricic* prevede l'incubazione delle librerie con i *Blocking Oligo*, ovvero degli oligonucleotidi complementari agli adattatori Solexa Illumina P5 e P7. Grazie alla presenza degli indici legati ai frammenti di DNA, è possibile aggiungere campioni diversi in una stessa reazione di cattura, in modo tale da utilizzare una quantità totale di DNA di 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 campione da aggiungere ad uno stesso *pool* viene preso in considerazione il valore di concentrazione ottenuto dal profilo *Agilent* nel range di lunghezza 150-400 bp; questo passaggio è fondamentale in quanto, al fine di ottimizzare la resa della cattura, è fondamentale che i campioni inseriti nello stesso *pool* siano fra loro equimolari.

7.4 Eluizione della libreria catturata

Trascorso il tempo di incubazione previsto dal protocollo *Maricic*, le catture sono state lavate in maniera da consentire la rimozione della di libreria non ibridata. A tal fine è stato utilizzato un *rack* magnetico (DynaMag™-Spin, Invitrogen) per separare le biglie che hanno legato il DNA dal resto della soluzione. terminate le fasi di lavaggio, le biglie con il DNA *target* adeso, sono state risospese in TET e trasferite in nuove eppendorf da 1,5 ml per procedere con l'eluizione della libreria catturata mediante l'utilizzo di NaOH che ha permesso la rottura dei legami a ponte d'idrogeno instauratisi tra i filamenti di sonda e di DNA *target*. A questo punto avvalendosi dell'utilizzo del magnete, le biglie sono state separate dal surnatante che contiene il DNA *target* di interesse, che viene trasferito in una nuova eppendorf. Successivamente, per evitare che il DNA possa subire un'ulteriore degradazione, la soluzione di eluizione alcalina è stata immediatamente neutralizzata aggiungendo una soluzione tampone contenente PBI (kit MinElute, Qiagen) e 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).

L'eventuale viraggio del colore verso il blu/violaceo è stato tamponato mediante l'aggiunta di ulteriore acido acetico al 20%. Gli eluati sono stati quindi purificati utilizzando il *MinElute Purification Kit* (Qiagen) e conservati a -20°C.

7.5 Quantificazione ed arricchimento del DNA catturato

Per verificare la resa della cattura è stata eseguita la *Real Time PCR* utilizzando *primers* complementari alle sequenze esterne agli indici (*indexing primers*), con il medesimo *kit* e protocollo precedentemente descritti. Dopo il successivo arricchimento del catturato, i campioni sono stati riuniti in un *pool* 2 nM e corsi sul sequenziatore *NGS MiSeq* utilizzando il *kit v3* e impostando 2x75 cicli di sequenziamento (*paired-end*).

8. Analisi dei dati di sequenza

Qualunque sia la piattaforma utilizzata per il sequenziamento, l'*output* è rappresentato da una serie di *file* in formato FASTQ che contengono tutte le sequenze (chiamate *reads*) generate. Per ciascun campione sequenziato con la tecnologia *MiSeq*, in particolare, vengono generati due *files* uno contenente le *reads forward* (R1) e uno contenente le *reads reverse* (R2) appartenenti alla medesima libreria.

Il file in formato FASTQ ha una struttura ben precisa mostrata di seguito:

```
@EAS139:136:FC706VJ:2:5:1000:12850 1:Y:18:ASEHGF
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATT
TGTTCAACT+
```


!"*(((***+))%%%++)(%%%%).1***-+*)"**55CCF>>>>>>>C

Si distinguono quattro tipologie di righe:

1. La prima riga inizia con il simbolo @ che precede il codice identificativo della sequenza;
2. La seconda riporta la sequenza nucleotidica;
3. La terza presenta il simbolo +, un separatore che segnala sia il termine della seconda riga, relativa alla sequenza, sia l'inizio della quarta riga (*Cock et al.*, 2010)
4. *Quality Score* (Q): sono riportati una serie di caratteri ASCII, dove ciascuno è relativo ad un nucleotide della sequenza ed è associato ad un determinato valore di qualità (*Illumina 2011*). Nel codice ASCII ogni carattere corrisponde ad un numero che è uguale al valore di qualità nel nucleotide a cui va sommato 33 (Q+33), per esempio, il carattere "I" è rappresentato nel codice ASCII dal numero "73" ed è associato ad un valore di qualità del nucleotide pari a 40 (40+33=73) (Tabella 1).

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 1: Codice ASCII; vengono riportati i caratteri ASCII con i relativi valori di qualità (Q) e probabilità (P) (Illumina, 2011)

La qualità relativa ad ogni base è espressa in PHRED *score* ed è legata alla probabilità P_e di richiamare una base errata dalla seguente equazione:

$$Q_{\text{PHRED}} = -10 \times \log_{10}(P_e)$$

Pertanto:

$$P_e = 10^{-Q/10}$$

Ad esempio nel caso in cui si abbia il valore di qualità $Q=40$, la probabilità di richiamare una base sbagliata sarà di 0,0001, con un'accuratezza del 99,99% (Tabella 2).

PHRED Quality Score	Probabilità di richiamare una base scorretta	Accuratezza nel richiamare una base corretta
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99,90%
40	1 in 10.000	99,99%
50	1 in 100.000	99,999%

Tabella 2: Valori di PHRED score associati alle relative probabilità di richiamare una base errata e corretta

In questo lavoro di tesi le analisi dei dati di sequenza sono state condotte utilizzando il pacchetto *FastQprocessing/MIA*; l'analisi del *pattern* di misincorporazione è stata eseguita con il *software Mapdamage* mentre l'autenticità delle sequenze è stata valutata con il *software ContamMix*.

8.1 Pacchetto FastQprocessing/MIA

Le *reads* generate dal sequenziamento sono state analizzate con il pacchetto di *script* denominati *FastQprocessing*, programmati in linguaggio *Python* (Kircher, 2012). Questi comandi consentono di estrapolare informazioni sulle *reads* quali per es. il numero totale, la lunghezza, la qualità del sequenziamento, il contenuto in GC ed il numero di nucleotidi non identificati (indicati con la lettera N).

8.1.1 Adapter trimming e merging delle reads

I primi *script* utilizzati hanno consentito di effettuare l'*adapter trimming*, ovvero l'eliminazione delle sequenze degli adattatori e parallelamente il *merging* delle *reads*, ovvero la sovrapposizione tra le *reads forward* e *reverse* (R1 e R2), ottenute dal sequenziamento. Per effettuare il *merging* è necessario impostare un diverso valore di sovrapposizione (*overlap*) a seconda che le *reads* abbiano una lunghezza di 75 bp o superiore. Questa strategia è stata adottata in quanto permette di evitare il recupero di eventuali sequenze contaminanti. Nel nostro caso, siccome il sequenziamento ha generato *reads* lunghe 76 bp, è stato impostato un parametro di *overlap* di 10 bp. Quando la percentuale di *reads* che hanno superato il *merging* è elevata vuol dire che il sequenziamento è di buona qualità e che quindi non si sono verificati errori nel sequenziamento e che i frammenti originali erano di brevi dimensioni: al massimo 142 bp ($(76+76) - 10 = 142$). Tutte le *reads* che hanno superato il *merging* sono state sottoposte al *quality filtering*, un filtro che ha permesso di eliminare le *reads* che presentavano un elevato contenuto in GC, tipiche del DNA batterico e quelle con più di cinque nucleotidi con un valore di PHRED *score* minore di 15. Successivamente le *reads merged*, cioè le sequenze di *reads* (R1 e R2) che si sono sovrapposte per almeno 10 bp e che hanno superato il *quality filtering*, sono state isolate da tutte le altre *reads* che non hanno risposto ai criteri impostati. Tale passaggio ha permesso che solo le *reads* che hanno superato i criteri impostati nel *merging* e nel *quality filtering* sono state utilizzate nella successiva fase di mappaggio.

8.1.2 Mapping assembly dell' mtDNA

Si è quindi proceduto al mappaggio delle *reads merged* e filtrate sul genoma mitocondriale utilizzando il *Mapping Iterative Assembler* (MIA) (Green et al. 2008). Tale programma è composto da una serie di *script* in grado di allineare tutte le *reads* appartenenti ad una determinata libreria, ad una sequenza di riferimento, creando una sequenza consenso. Questo programma è stato appositamente studiato per il DNA antico: infatti per l'allineamento e la creazione della sequenza consenso viene utilizzata una matrice di sostituzione posizione-specifica che tiene conto dei possibili danni presenti sul DNA antico. Il primo *script* MIA è in grado di collassare tutte le copie di una stessa sequenza in un'unica *reads* consenso richiamando la base con il più alto valore di qualità in caso di *mismatch*. Tale passaggio è stato effettuato dal momento che il flusso di lavoro stabilito per il DNA antico prevede vari passaggi di arricchimento/amplificazione per aumentare la quantità di DNA endogeno pertanto, per ogni campione, si sono generate molteplici copie identiche di uno stesso frammento di DNA. L'*output* dello *script* è rappresentato da due o più *files* di mappaggio *unique* (con estensione .maln) contenenti le *reads* copie della stessa sequenza collassate e da due o più *files non-unique* (con estensione .maln), nei quali le *reads* clonali sono state mappate mantenendole separate. I file *unique* e *not-unique* sono stati poi utilizzati come *input* per i successivi *script* per generare la sequenza consenso mappando 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* ha permesso di definire il *cluster factor*, un

valore che ha fornito una stima della profondità del sequenziamento di ciascun campione. Lo script *MapAssembler* ha processato le *reads unique* fornendo una serie di informazioni sul mappaggio, come l'*average coverage*, ovvero la profondità del sequenziamento per ogni singola posizione (quante volte ciascuna base è stata richiamata in quella posizione nelle *reads*), l'*average length*, ovvero la lunghezza media dei frammenti e il *fold coverage*, cioè quante basi sul totale delle 16.569 del mtDNA sono state coperte almeno 1, 2, 3, 4 e 5 volte. Quest'ultimo dato ci ha permesso di conoscere la percentuale di copertura dell'mtDNA e se ci sono state o meno posizioni non sequenziate.

8.4 Analisi del *pattern* di misincorporazione

L'autenticità delle sequenze di DNA mitocondriale ottenute è stata valutata utilizzando strumenti bioinformatici che vanno a stimare il *pattern* di misincorporazione. Le *reads merged* che hanno superato il *quality filtering* sono state mappate sulla sequenza di riferimento mitocondriale con BWA (*Burrows - Wheeler Aligner*; Li and Durbin, 2009) che ha restituito come *output* un file con estensione .bam. Questo *file* è stato utilizzato come *input* del programma *MapDamage 2.0*, un *framework* scritto in Python e R, in grado di rilevare e quantificare i *pattern* di danneggiamento all'interno dei dati di sequenza di aDNA generati dalle piattaforme NGS (Ginolhac et al. 2011; Jönsson et al. 2013). Nei frammenti di DNA antico si registra, solitamente a causa dei processi di deaminazione *post-mortem* che coinvolgono la citosina, un aumento del tasso di sostituzione C→T al 5' della *read* e un

corrispondente aumento della sostituzione G→A al 3' (Briggs et al. 2007). *MapDamage 2.0* si fonda sull'assunzione che gli eventi di danneggiamento dipendano solo dalla posizione sulla sequenza e dalla deaminazione *post-mortem*, riprendendo il modello statistico descritto da Briggs et al. (Briggs et al. 2007). Questo modello usa come fattori chiave le rotture del filamento, le estremità sporgenti e la deaminazione della citosina. *MapDamage* permette pertanto di effettuare una stima di tipo *bayesiano* che prende in considerazione quattro parametri fondamentali:

- **Lambda (λ):** definisce la probabilità di terminazione di una sporgenza;
- **DeltaD (δ_D):** definisce la probabilità di deaminazione della citosina all'interno del doppio filamento;
- **DeltaS (δ_S):** indica la probabilità di deaminazione della citosina all'interno del singolo filamento;
- **Theta (θ):** definisce il tasso di differenza media tra la sequenza di riferimento e quella del campione sequenziato non causato da danni al DNA.

Il *software* è in grado di valutare i *patterns* di sostituzione nucleotidica andando ad allineare la sequenza di riferimento a quella del campione. In seguito al *merging* vengono valutate le sostituzioni nucleotidiche sia alle estremità 5' e 3' che all'interno della sequenza campione. La lettura viene eseguita in base al parametro “-a” impostato, che indica quante

posizioni nucleotidiche devono essere rappresentate. In Figura 11 viene riportato un esempio di come lavora il *software*.

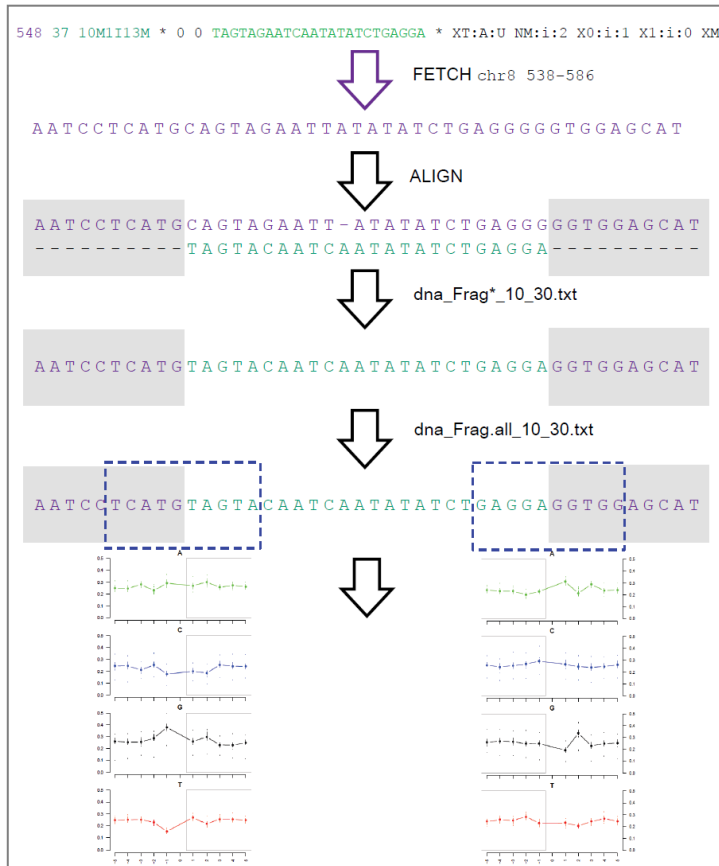


Figura 11: Principio di funzionamento del *software mapDamage*.

In viola viene rappresentata la sequenza di riferimento ed in verde quella del campione in esame. Queste due sequenze vengono allineate ottenendo un unico frammento con all'estremità le basi nucleotidiche della sequenza di riferimento (in viola) e internamente (in verde) le basi

della sequenza campione. Impostando il valore di “-a” a cinque nucleotidi, il *software* delimita una regione di lettura (rettangolo blu) di cinque nucleotidi, considerati a partire dalla sequenza di riferimento a quella del frammento campione sequenziato. Quindi in totale per ogni estremità vengono considerati 10 nucleotidi di cui 5 della sequenza di riferimento e 5 della sequenza campione. In base a quanto registrato *MapDamage* origina dei *plot*, denominati *Fragmisincorporation* ciascuno per ogni base nucleotidica, in cui vengono rappresentate, in maniera grafica, le frequenze di sostituzione. I grafici riportano, sull'asse delle ordinate, il valore della frequenza di sostituzione e, sull'asse delle ascisse, il numero di nucleotidi impostati per la valutazione. Queste posizioni nucleotidiche vengono riportate sia con valori negativi che con valori positivi andando ad indicare con i primi, le posizioni nucleotidiche della sequenza di riferimento e con i secondi, le posizioni nucleotidiche della sequenza del campione. Quest'assegnazione è valida se si considera il filamento 5'; se invece viene considerato il filamento 3', le posizioni vengono riportate con valori negativi per la sequenza campione e con valori positivi per la sequenza di riferimento.

8.5 Analisi delle contaminazioni

8.5.1 Software ContamMix

L'analisi delle contaminazioni è stata effettuata utilizzando il *software ContamMix* il quale, mediante un approccio di tipo *bayesiano* (Fu et al. 2013), calcola la probabilità che il profilo di DNA ottenuto da ciascun campione derivi da un unico individuo. Il *software* utilizza i *file .bam*

generati con BWA e gli allineamenti in formato .fasta e calcola questa probabilità sulla base di due *file*:

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

Il *software* permette di stimare la frazione di *reads* che mappano con un punteggio più elevato sulla sequenza consenso ricostruita, piuttosto che su quelle dei potenziali contaminanti con un intervallo di confidenza del 95%. Il *software* considera le n *reads* del campione, la sequenza consenso del campione e gli m genomi mitocondriali moderni ($m= 311$) tenendo in considerazione il fatto che le *reads* siano state ottenute da $m+1$ genomi. Per ogni *read* $i \in \{1, \dots, n\}$ e per ogni genoma $j \in \{1, \dots, m\}$, dove $j=0$ rappresenta la sequenza consenso del campione, si definiscono tre parametri:

- M_{ij} : numero di basi della *read* i presenti nel genoma j (dimostra che la *read* i appartiene al genoma j)
- N_{ij} : numero di basi della *read* i non presenti nel genoma j (dimostra che la *read* i non appartiene al genoma j)
- I_{ij} : inserzioni e delezioni della *read* i presenti (1) o non presenti (0) nel genoma.

Il *software* è in grado di scartare le *reads* che contengono quelle inserzioni e delezioni che non sono presenti in nessun genoma, assumendo che siano state generate da errori durante il processamento delle sequenze. La maggior parte delle informazioni che riguardano il tasso di errore ϵ , cioè la probabilità che per una singola base avvenga un errore di sequenziamento, sono state ottenute dalle regioni di mtDNA in cui tutti gli $m+1$ genomi sono risultati identici; il tasso con cui le *reads* differiscono da questa sequenza fissa è il tasso di errore. Dato il tasso di errore, viene usato un algoritmo MCMC (*Monte Carlo Markov Chain*) per stimare la proporzione reale sotto un modello più generale in cui la contaminazione può derivare da un numero arbitrario di individui ($m > 1$).

9. Determinazione dell'aplogruppo mitocondriale

Gli aplogruppi mitocondriali sono definiti come un raggruppamento di aplotipi i quali, a loro volta, sono definiti a partire dalla somma dei siti che presentano variazioni nella sequenza nucleotidica. Per la determinazione degli aplogruppi è stato utilizzato il *software online HaploFind*. Si tratta di una piattaforma bioinformatica che sfrutta un algoritmo in grado di classificare in modo semplice ed efficace nuove sequenze di DNA mitocondriale combinando le informazioni ottenute da *PhyloTree* (Vianello *et al.* 2013) con quelle acquisite analizzando le mutazioni conservate di migliaia di sequenze mitocondriali recuperate in *GenBank*. Il programma confronta la sequenza di riferimento *rCRS* del genoma mitocondriale con quella del campione in esame individuando i polimorfismi ed assegnando l'aplogruppo corrispondente di appartenenza. Il *software* attribuisce anche un *score*, cioè un punteggio

che esprime la probabilità che l'aplogruppo assegnato a ciascun campione sia effettivamente quello di appartenenza.

RISULTATI E DISCUSSIONE

I campioni ossei ed il dente, previa pulizia e polverizzazione, sono stati sottoposti ad estrazione del DNA seguendo il protocollo messo a punto da *Dabney (Dabney et al. 2013)* ottimizzato per il recupero di frammenti di DNA molto corti provenienti da materiale altamente degradato. Una volta estratti, i campioni sono stati quantificati utilizzando il *Quantifiler® Trio DNA Quantification Kit*. I risultati di quantificazione sono riportati nella Tabella 3.

Campioni	Concentrazione DNA (pg/μl)		
	LAT	SAT	YT
C-33	0	2.9	2.8
C-34	6.1	117	107
C-21	0	28.1	22.9
C-23	4.3	32,7	25.4
C-32	NON DETERMINATO		
C-35	1.2	17.8	9.3

Tabella 3: Risultati della quantificazione degli estratti con il *Quantifiler Trio DNA Quantification Kit*. I valori di concentrazione per ciascun campione, espressi in pg/μl, sono relativi alle regioni *target* selezionate: LAT=*Large Autosomal*, SAT=*Small Autosomal* e YT=*Y-Chromosome*.

Come si può osservare dai risultati riportati in tabella 3, non è stato possibile ottenere alcuna quantificazione per il campione C-32 a causa, presumibilmente, della scarsa quantità del frammento osseo di partenza e

del materiale genetico presente nell'estratto. A causa inoltre dell'elevata degradazione del DNA, nei campioni C-33 e C-21, non è stato possibile quantificare il *target* relativo alla porzione autosomica maggiore (*Large Autosomal*). Tutti i campioni presentavano invece una quantificazione sia del frammento piccolo autosomico (*Small Autosomal*) che del *target* del cromosoma Y. La presenza di quantificato nel frammento del cromosoma Y in tutti i campioni analizzati ci ha portato ad ipotizzare di essere in presenza di individui di sesso maschile.

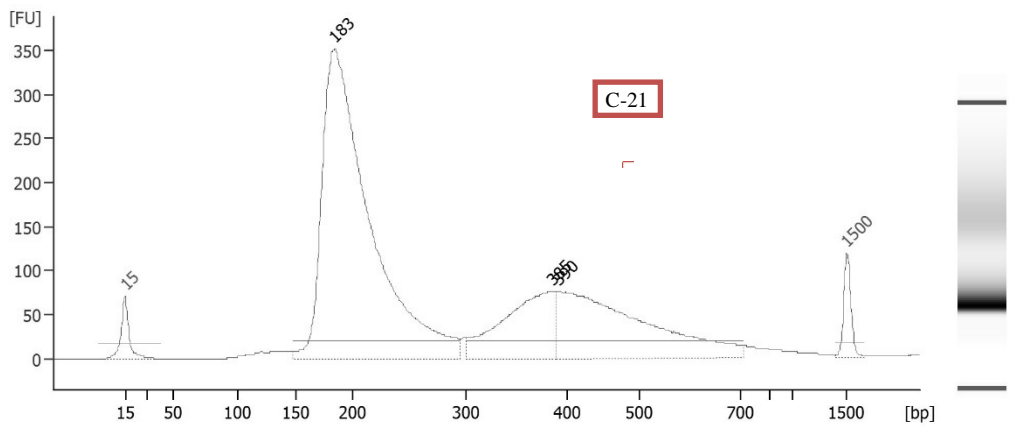
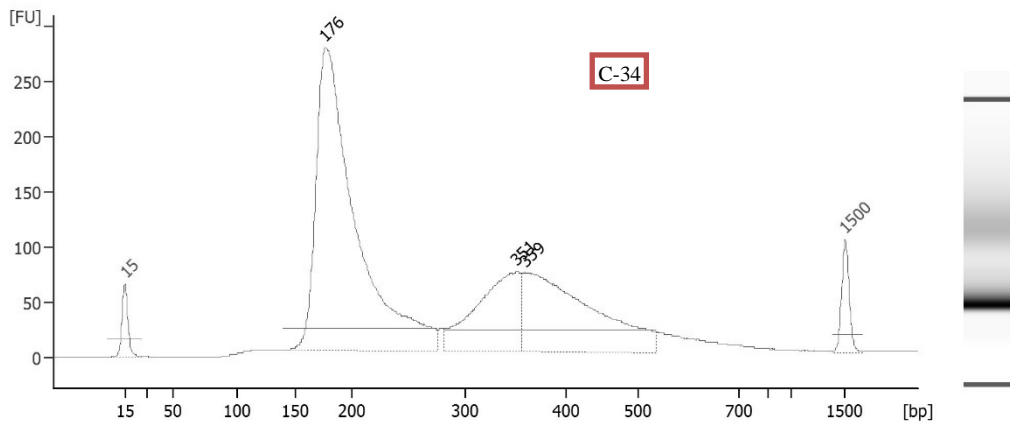
Successivamente sono state allestite le librerie Illumina secondo il protocollo di *Meyer e Kircher*. Una volta ottenute le librerie, queste sono state indicizzate mediante *indexing* PCR; per ciascun campione e controllo negativo (K- estrazione e K- library) è stata scelta una combinazione di coppie di indici diversa con lo scopo di poterli identificare in maniera univoca al momento del sequenziamento.

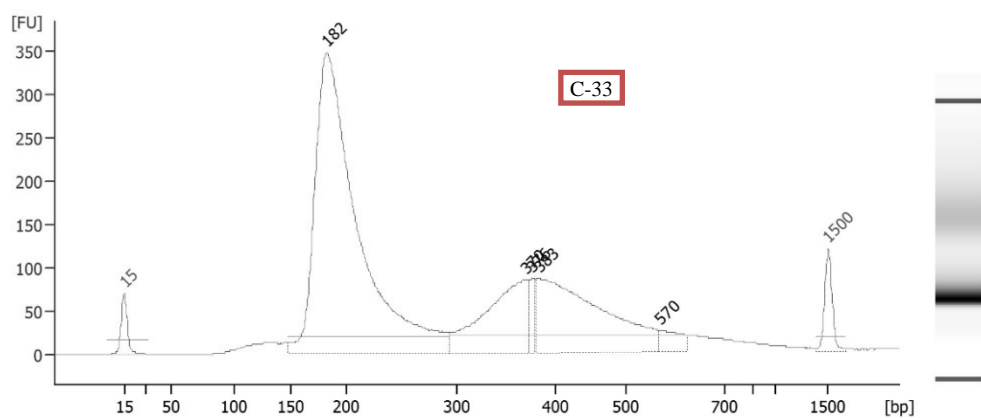
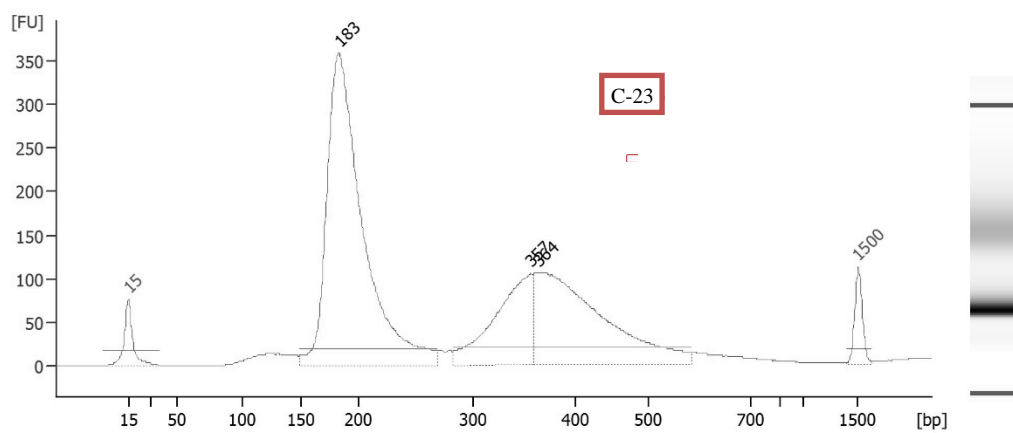
Al termine della fase di indicizzazione, è stata eseguita la *RT-PCR*, una PCR che ha permesso di quantificare la concentrazione di DNA presente. I risultati della quantificazione sono riportati in Tabella 4 in cui, nella prima colonna, è riportato il nome del campione analizzato, nella seconda il numero di copie/ μ l di ciascun campione ed infine nella terza la concentrazione totale del DNA calcolata moltiplicando il numero di copie/ μ l per il volume di reazione, nel nostro caso 48 μ l.

Campione	N° copie/ μ l	N° copie di DNA in 48 μ l
C-33	1,30E+12	6,25E+13
C-34	1,31E+12	6,30E+13
C-21	1,97E+12	9,45E+13
C-23	2,25E+12	1,08E+14
C-32	1,32E+12	6,34E+13
C-35	8,79E+11	4,22E+13

Tabella 4: Risultati della *RT-PCR post-indexing*. Viene riportato il numero di copie di partenza/ μ l e il numero di copie di DNA nel volume totale di reazione

Una volta quantificati, i campioni sono stati arricchiti mediante l'utilizzo del *kit Herculase* e la quantità e qualità dei frammenti sono state valutate mediante l'utilizzo dell'*Agilent BioAnalyzer2100*. I profili ottenuti per ciascun campione sono riportati in Figura 12:





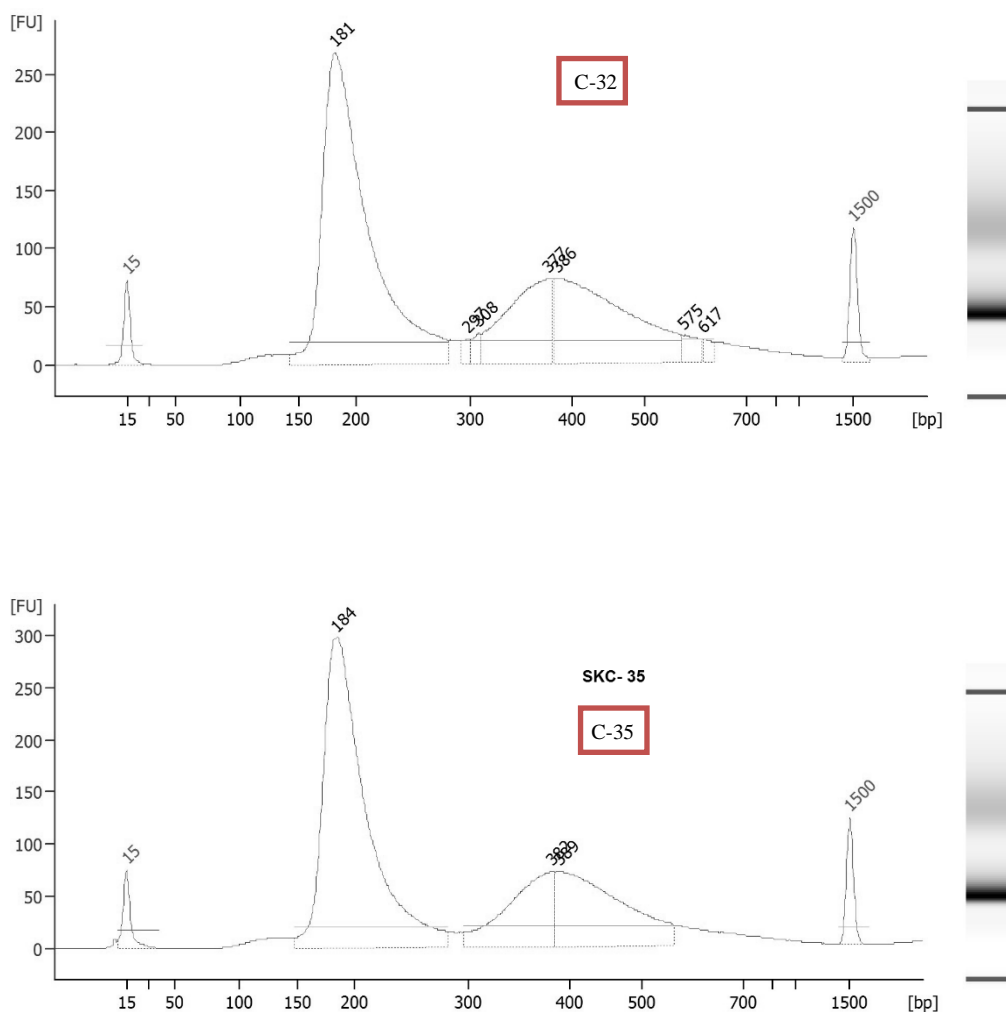


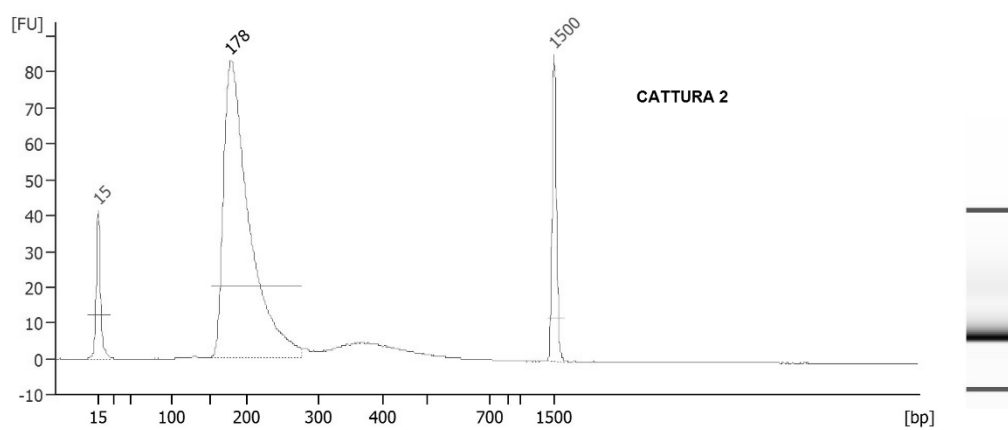
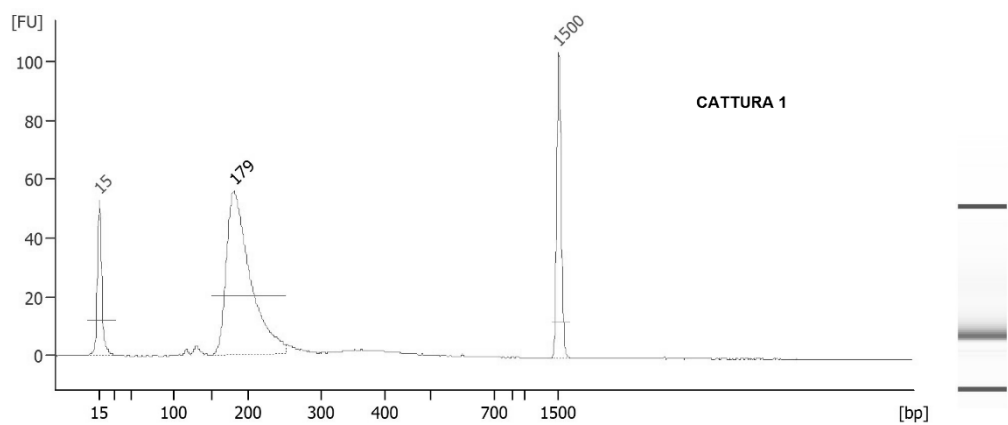
Figura 12: Risultato dell'elettroforesi quantitativa effettuata mediante *Agilent BioAnalyzer2100*. Sull'asse delle Y è riportata la concentrazione del DNA espressa in unità di fluorescenza (FU) mentre sull'asse delle X la lunghezza dei frammenti in paia basi. I picchi a 15 e 1500 bp corrispondono rispettivamente al picco minore e al picco maggiore del marcatore a concentrazione nota e di lunghezza (in bp) nota.

Quando si lavora con campioni antichi il *range* di lunghezza dei frammenti che mostrano una maggiore concentrazione è compresa tra le 170 e le 350 bp, con una media vicina alle 200 bp compatibile con la lunghezza dei frammenti di DNA a cui sono stati attaccati gli adattatori e gli indici ad entrambe le estremità. Dall'osservazione dei grafici in figura 12 appare subito evidente come tutti i campioni mostravano un picco principale, tipico del DNA antico, intorno alle 180 bp ed un ulteriore picco, più basso e più slargato, ad una lunghezza maggiore, compresa tra le 300 e le 700 bp che potrebbe essere associato alla presenza di DNA contaminante moderno o alla formazione di *heteroduplex*, molecole ibride generate dall'appaiamento di filamenti complementari appartenenti a frammenti diversi. Dopo aver valutato la qualità del profilo di quantificazione ottenuto è stata stimata la concentrazione del DNA in un *range* compreso tra le 150 e 400 bp. Successivamente si è proceduto alla cattura del DNA mitocondriale di interesse mediante protocollo *Maricic et al.* (*Maricic et al.* 2010). Una volta catturato il DNA mitocondriale di interesse, è stata effettuata una quantificazione mediante *RT-PCR* ed il risultato riportato in Tabella 5 mostra, come atteso, una notevole riduzione della concentrazione del DNA catturato rispetto alla concentrazione delle librerie grazie alla selezione/cattura solo dei frammenti di DNA mitocondriale umano di interesse.

Campione	Copie post-cattura/μl	Copie totali post-cattura
Cattura 1	5,86E+05	5,86E+06
Cattura 2	1,36E+06	1,36E+07
Cattura 3	7,82E+05	7,82E+06
Cattura 4	7,02E+05	7,02E+06

Tabella 5: Risultati *RT-PCR* post-cattura. La colonna 1 riporta il nome dei campioni (ciascuna cattura contiene due campioni, due librerie che sono state catturate insieme); La colonna 2 riporta il numero di copie per μ l, mentre la colonna 3 riporta il numero di copie totali ottenuto dal prodotto tra le copie post-cattura/ μ l ed il volume di reazione.

Un arricchimento del catturato è stato ottenuto mediante amplificazione con il *kit Herculanase*. La valutazione quali-quantitativa delle librerie catturate ed amplificate è stata effettuata mediante *Agilent BioAnalyzer 2100* ed i risultati sono riportati in Figura 13:



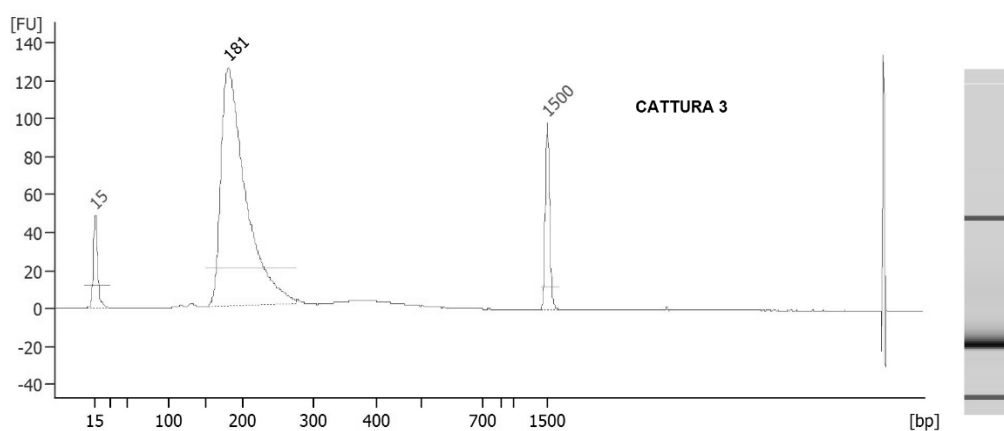


Figura 13: Risultati dell'amplificazione post-cattura all'Agilent BioAnalyzer2100.

Tutti i profili riportati in figura 13 mostravano la presenza, come atteso, di un unico picco fra le 178 e le 180 bp. I catturati, riuniti in un unico *pool* equimolare pari a 2nM, sono stati sequenziati. Il sequenziamento ha fornito un *output* costituito dalle *reads* R1 (*forward*) e R2 (*reverse*) per ciascun campione che sono state poi sottoposte al processo di *demultiplexing*, cioè separate sulla base degli indici assegnati. Le sequenze ottenute sono state quindi analizzate mediante il ricorso ad una *pipeline* specifica per il DNA antico: *FastQprocessing/MIA*. Il pacchetto *FastQprocessing* permette di effettuare contemporaneamente l'*adapter trimming*, ovvero l'eliminazione delle sequenze relative agli adattatori ed il *merging*, la fusione fra la *read* R1 e la *read* R2 con una regione di sovrapposizione (*overlapping*) di 10 bp. Questo pacchetto di *script* ha

generato informazioni riguardanti il numero originario di *reads* R1 e R2, il numero totale di *reads* (*original number of reads*) dato dalla somma delle *reads* di entrambi i filamenti, le *reads* che non hanno superato il *merging* (*total #separated reads*) in quanto non conformi ai parametri di *overlapping* impostati ed infine la percentuale di *merging* definita dal rapporto tra le *merged reads* e le *reads* totali:

$$\frac{\text{MERGED READS}}{\text{(MERGED READS + TOTAL \#SEPARATED READS)}}$$

I dati ottenuti dall'analisi sono riportati in Tabella 6.

Sample	original number of reads	Merged reads	Forward reads	Reverse reads	Total #separated reads	% of merged
C-33	1923842	942218	19545	19545	39090	96.01
C-34	1481500	725657	14969	14969	29938	96.03
C-21	1703404	833578	17622	17622	35244	95.94
C-23	1237746	605987	12110	12110	24220	96.15
C-32	1374230	671961	14807	14807	29614	95.77
C-35	1082018	531862	8927	8927	17854	96.75

Tabella 6: Risultati del *merging*.

Come si evince dalla tabella sopra riportata il campione C-33 presentava il numero totale di *reads* più elevato (1.923.842), seguito dal campione C-21 con 1.703.404 *reads* totali, C-34 e C-32 con rispettivamente 1.481.500 e 1.374.230 *reads* ed infine i campioni C-23 e C-35 con un

minor numero di *reads* (rispettivamente 1.237.746 e 1.082.018 *reads*). Per quanto riguarda la percentuale di *merging*, tutti i campioni hanno mostrato valori che oscillavano fra il 95 e il 96% rientrando, seppur con qualche lieve deviazione, nel *range* ottimale, compreso fra il 61,85% e il 95,4%, stabilito per i frammenti di piccole dimensioni (Kircher 2012). Le *reads merged* sono state quindi sottoposte al *quality filtering* al fine di scartare tutte le *reads* che non rispondevano ai criteri di qualità stabiliti per il DNA antico. I risultati sono riportati in Tabella 7:

Sample	Merged reads	Total separated reads	Total quality filtered	% merged quality filtered
C-33	939589	27504	967093	97,16
C-34	723768	22742	746510	96,95
C-21	831175	26493	857668	96,91
C-23	605009	9257	614266	98,49
C-32	669874	20982	690856	96,96
C-35	530210	12194	542404	97,75

Tabella 7: Risultati relativi alle *merged reads* sottoposte a *quality filtering*.

In colonna 2 sono riportate le *reads merged* che hanno superato il *quality filtering* mentre nella colonna 3 le *total separated reads* che rappresentano quelle sequenze che, non presentando i requisiti richiesti, sono state scartate. Dal rapporto fra le *merged reads* e le *merged reads + total separated reads* è stata calcolata la percentuale di *merged quality filtered* (colonna 4). Come si evince dai risultati riportati in colonna 4, tutti i campioni presentavano una percentuale di *merged quality filtered*

reads superiore al 96%, valore questo che ha messo in evidenza la buona riuscita del sequenziamento dal momento che la maggior parte delle sequenze che hanno superato il *merging* hanno superato anche il *quality filtering*. Successivamente, le *reads merged* che hanno superato il *quality filtering* sono state mappate sulla sequenza mitocondriale di riferimento *rCRS* mediante il *mapping assembly* utilizzando il pacchetto di *script MIA*, specifico per il DNA degradato. Il primo *script* che è stato usato ha permesso di collassare, in un'unica sequenza, tutte le *reads* clonali allineate con la medesima direzione e con le stesse coordinate di inizio e fine, generando frammenti unici ("*fragments unique*", Tabella 8 colonna 3) e di allineare tutti i frammenti separati e quindi non unici ("*fragments not unique*", Tabella 8 colonna 2). Il rapporto fra sequenze *not unique* e *unique* ha generato il *cluster factor* (Tabella 8 colonna 5), un valore che indica il numero di copie di un frammento che sono state in media sequenziate ed è in grado di fornire una stima del grado di profondità del sequenziamento. Valori di *cluster factor* compresi fra 2 e 10 vengono considerati ottimali, mentre invece valori al di sotto di questo intervallo sono indice di un sequenziamento poco profondo che non ha permesso l'adeguato sequenziamento dei frammenti catturati. Al contrario valori superiori a 10 indicano un sequenziamento troppo profondo che potrebbe produrre informazioni ridondanti. I valori di *cluster factor* ottenuti per i campioni analizzati rientrano pienamente nel *range* ottimale, fatta eccezione dei campioni C-34 e C-21 che hanno mostrato valori superiori.

Sample	Fnu	Fu	F (filter)	CF	AC	AL
C-33	87547	15432	15432	5.67	55.387	59.46
C-34	163565	10248	10248	15.96	40.824	66.00
C-21	96779	8933	8933	10.83	37.345	69.26
C-23	67566	19619	19619	3.44	69.859	58.99
C-32	2328	253	253	9.20	0.910	59.61
C-35	2701	817	817	3.30	3.435	69.66

Tabella 8: Risultati del mappaggio. Fnu=*fragment (not unique)*; Fu= *fragment (unique)*; CF=*cluster factor*; AV=*average coverage*; AL=*average lenght*.

In tabella 8 vengono riportati anche i valori relativi alla copertura media (*average coverage*) cioè il grado di profondità del sequenziamento dell'intero genoma mitocondriale. L'*average coverage* è definito in base alla presenza di frammenti unici con coordinate diverse che si sovrappongono in una determinata regione del genoma; pertanto, più tale valore è elevato, maggiore sarà l'attendibilità della sequenza consenso ricostruita per ciascun campione. La copertura media è risultata elevata per i campioni C-23, C-33, C-34 e C-21 (rispettivamente 69.8, 55.3, 40.8 e 37.3), mentre i campioni C-32 e C-35 hanno mostrato una copertura media estremamente bassa, rispettivamente 0.9 e 3.4. Nell'ultima colonna della tabella 8 è stata infine riportata la lunghezza media dei frammenti analizzati (*average lenght*) che si aggira, per tutti i campioni, intorno a valori compresi fra le 58 e 69 bp. Tali valori rientrano pienamente nel *range* di 40-150 bp stabilito per il DNA antico. Infine l'ultimo *script MIA* ha permesso di determinare la copertura di ciascun

nucleotide della sequenza di DNA mitocondriale fino al quinto livello di profondità (5 *fold*) (Tabella 9).

Sample	Copertura 1 <i>fold</i>	Copertura 2 <i>fold</i>	Copertura 3 <i>fold</i>	Copertura 4 <i>fold</i>	Copertura 5 <i>fold</i>
C-33	99.97 %	99.97 %	99.90 %	99.69 %	99.46 %
C-34	100.00 %	100.00 %	99.92 %	99.90 %	99.87 %
C-21	100.00 %	100.00 %	100.00 %	100.00 %	99.96 %
C-23	100.00 %	100.00 %	100.00 %	99.99 %	99.96 %
C-32	57.45 %	23.13 %	7.58 %	1.77 %	.52 %
C-35	94.37 %	81.14 %	64.67 %	44.61 %	28.04 %

Tabella 9: Percentuale di copertura del DNA mitocondriale ai vari *fold*.

Dalla tabella si evince che i campioni C-33, C-34, C-21 e C-23 presentavano una copertura pari al 99% fino al 5 *fold*. Il campione C-35 invece presentava una copertura elevata all'1 *fold* che tendeva però a diminuire con l'aumentare del livello di copertura (si passa dal 94.37% all'1 *fold* al 28.04% al 5 *fold*); il campione C-32 infine presentava la percentuale di copertura più bassa, basti osservare che all'1 *fold* la percentuale di copertura risultava essere pari al 57.45%: solo poco più della metà della sequenza mitocondriale è risultata coperta almeno una volta. Al fine di valutare l'autenticità del dato genetico ottenuto, il *pattern* di danneggiamento a cui normalmente va incontro la doppia elica antica è stato stimato mediante *MapDamage 2.0*. Questo *software* permette di determinare il *pattern* di danneggiamento *post-mortem* dovuto ad alterazioni a carico delle molecole del DNA fornendo

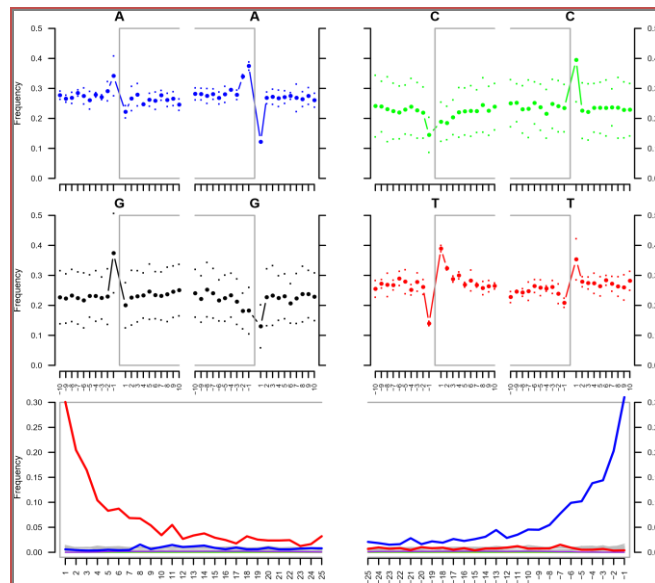
informazioni sul grado di conservazione del materiale genetico. Dal momento che i fenomeni degradativi portano alla formazione di frammenti con sporgenze all'estremità, è evidente che la frequenza degli eventi di sostituzione C→T e G→A sia maggiore a livello delle estremità di una molecola rispetto ad una porzione interna a doppio filamento. Dal rapporto tra il numero di C→T e il numero totale di C all'estremità 5' e tra G→A e G all'estremità 3' per una determinata posizione, sono state calcolate le frequenze di ciascuno dei due tipi di misincorporazione. Il programma *mapDamage* calcola i *pattern* di misincorporazione fino alla venticinquesima posizione ma in Tabella 10, per praticità, è stata riportata solo la prima posizione.

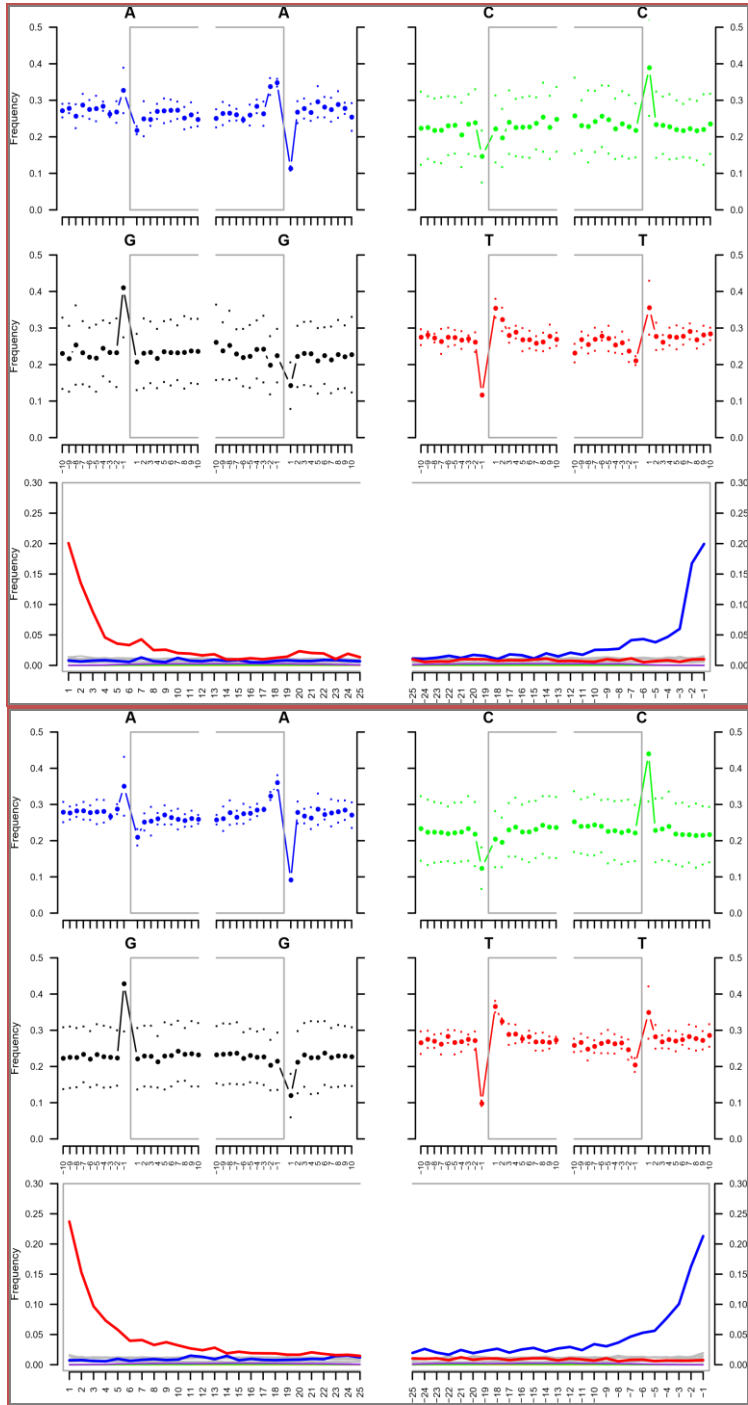
Sample	Map Damage	
	5' C>T %	3' G>A %
C-33	0.301	0.31
C-34	0.201	0.199
C-21	0.237	0.213
C-23	0.242	0.221
C-32	0.033	0.041
C-35	0.154	0.114

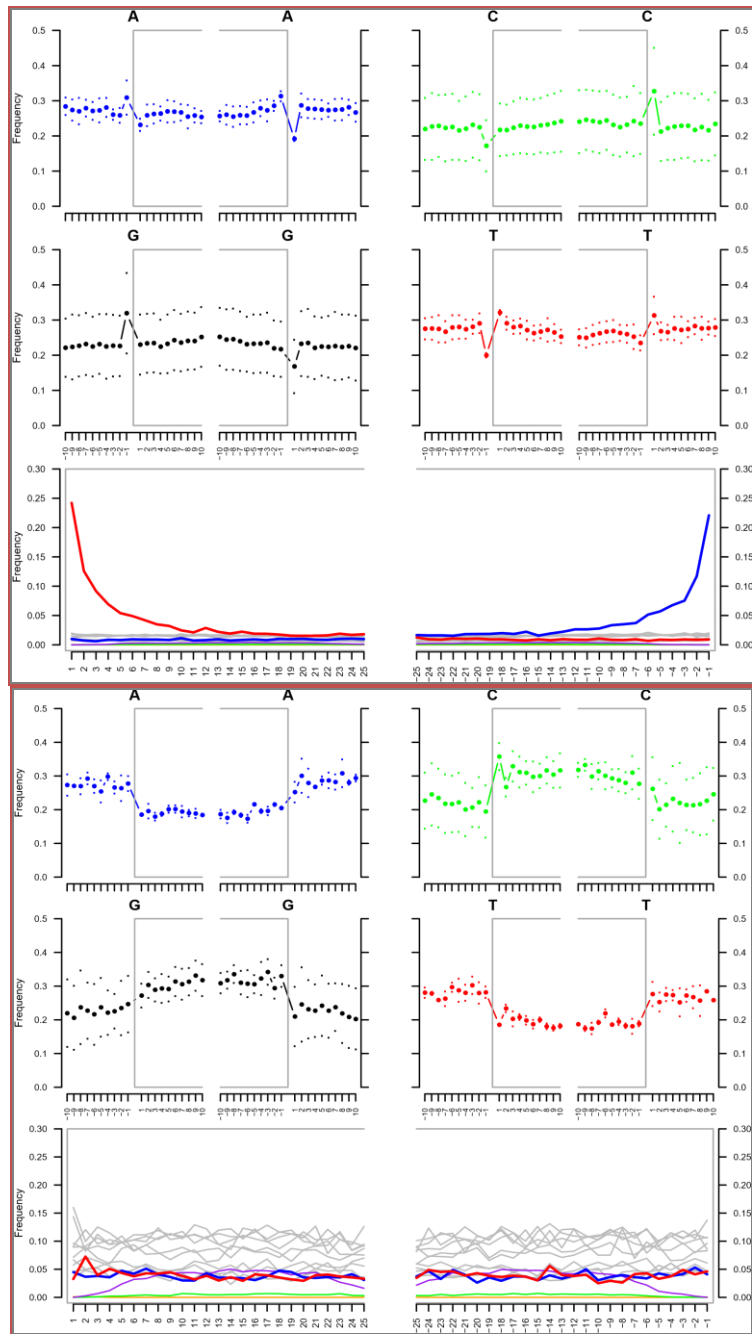
Tabella 10: Confronto dei *pattern* di misincorporazione alle estremità 3' e 5' relativi alla prima posizione.

I risultati riportati in tabella mostrano valori di *pattern* di misincorporazione tipici del DNA antico almeno per cinque dei sei

campioni analizzati. È stato infatti dimostrato che la frequenza di sostituzioni C→T è correlata con l'età del reperto e che per i campioni antichi non deve essere inferiore a 0.15- 0.25 (Sawyer *et al.* 2012). Il campione C-32 presenta invece una frequenza di misincorporazione al 5' e al 3' inferiore ai valori stabiliti per il DNA antico facendo questo supporre che il DNA estratto da questo campione possa essere per lo più DNA moderno oppure che, data la bassa copertura di DNA mitocondriale ottenuta, il calcolo della frequenza di sostituzione effettuato possa non risultare indicativo di una situazione reale. I risultati ottenuti dal *software* sono stati presentati sotto forma di grafici, i *fragmisincorporation plot* (Figura 14) che mettono in evidenza le frequenze di misincorporazione per ciascuna base nucleotidica (i quattro mini *plot* in alto) e l'andamento delle frequenze delle sostituzioni C>T e G>A (rispettivamente linea rossa e blu) nei due mini *plot* riportati in basso nella Figura.







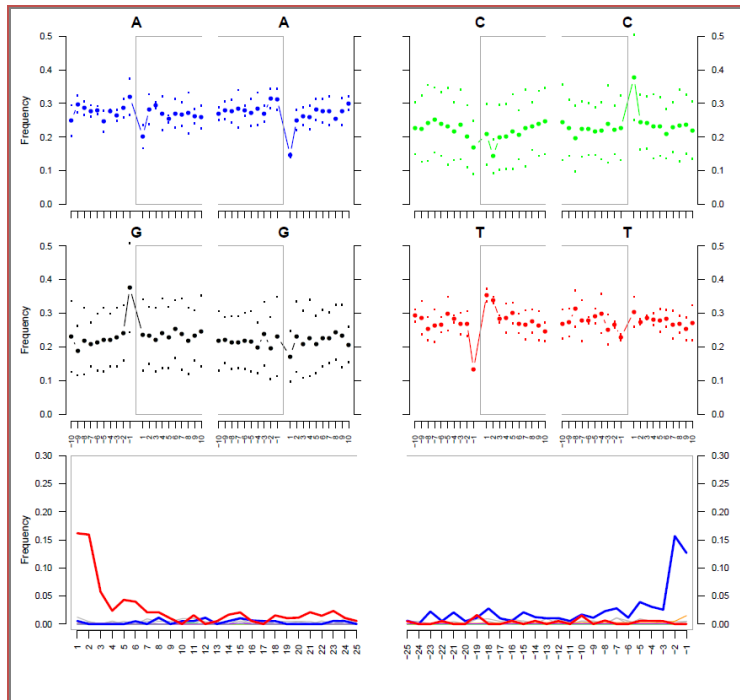


Figura 14: *Fragmisincorporation plot* relativi, in successione, ai campioni C-33, C-34, C-21, C-23, C-32 e C-35.

I grafici riportano sull'asse delle ordinate il valore della frequenza di sostituzione e sull'asse delle ascisse il numero di nucleotidi impostati per la valutazione. Come già accennato nel paragrafo dedicato al *software*, queste posizioni nucleotidiche vengono riportate sia con valori negativi che con valori positivi andando ad indicare, con i primi, le posizioni nucleotidiche della sequenza di riferimento e con i secondi, le posizioni nucleotidiche della sequenza del campione. Quest'assegnazione è valida se si considera il filamento 5'; se invece viene considerato il filamento 3' le posizioni vengono riportate con valori negativi per la sequenza

campione e con valori positivi per la sequenza di riferimento. Nel nostro caso il valore -a impostato è stato di 10 nucleotidi, per cui nell'asse delle ascisse sono riportate le 10 posizioni nucleotidiche relative alla sequenza di riferimento e alla sequenza consenso. I “confini” delle due sequenze vengono delimitate da una linea di colore grigio. Per ciascuna base vengono prodotti due grafici che indicano le frequenze di sostituzione C>T valutate all'estremità 5' e G>A valutate all'estremità 3', riportate, nei grafici, rispettivamente con una linea rossa e con una linea blu. Dall'andamento dei dati registrati nei *plot* relativi al campione C-33, C-34, C-21 e C-23 è stato possibile osservare un andamento omogeneo ad entrambe le estremità con una frequenza di sostituzione C>T (linea rossa) al 5' elevata che raggiunge valori di circa 0.20-0.30, mentre la frequenza di sostituzione G>A (linea blu) è praticamente 0 ed una frequenza di sostituzione al 3' G>A di circa 0.20-0.30, mentre la frequenza di sostituzione C>T è prossima allo 0. Le frequenze di sostituzione C>T e G>A hanno mostrato invece un andamento completamente diverso nei campioni C-32 e C-35 (ultimi due grafici nella Figura 14). In entrambi i casi le linee presentavano un profilo pressoché piatto, differente da quello che normalmente si osserva in un campione antico, con valori di frequenza compresi tra 0.05 e 0.1. Inoltre in entrambi i campioni, i *mini plot* hanno messo in evidenza altre possibili mutazioni, riportate in grigio, viola e verde. La linea arancione che compare nei grafici rappresenta invece il *soft-clipping*, ovvero le basi situate alle estremità delle *reads* che non essendo state allineate alla sequenza di riferimento poiché differenti, non sono state prese in considerazione nell'analisi.

I campioni sono stati successivamente sottoposti ad un'analisi *bayesana* (*Contaminaton Check*) per valutare la probabilità che il profilo di DNA ottenuto da ciascun campione appartenga ad un singolo individuo mediante il *software ContamMix* che ha preso in considerazione l'allineamento delle *reads* alla sequenza consenso e l'allineamento della sequenza consenso del campione alle 311 sequenze di DNA mitocondriale che rappresentano potenziali contaminanti. In questo modo è stato stimato il numero di *reads* che mappavano con un punteggio più elevato sulla sequenza consenso piuttosto che su quelle dei potenziali contaminanti con un intervallo di confidenza del 95%. I risultati ottenuti sono riportati in Tabella 11.

<i>Sample</i>	<i>Map authentic</i>
C-33	0.912304
C-34	0.903813
C-21	0.954469
C-23	0.9826951
C-32	0.9988368
C-35	0,9777565

Tabella 11: Valori di *map authentic* ottenuti mediante utilizzo del *software ContamMix*.

Dai risultati ottenuti si osserva che tutti i campioni mostravano valori di *map authentic* superiori al 90% facendo questo ipotizzare di essere in presenza di un profilo proveniente da un unico individuo.

Associando i valori del *map authentic* riportati in Tabella 11 a quelli delle frequenze di misincorporazione ottenute dal *software MapDamage* (Tabella 10) è stato possibile ipotizzare di essere in presenza di profili unici, provenienti cioè da un unico individuo, ed autentici cioè profili endogeni “antichi” per i campioni C-33, C-34, C-21, C-23 e C-35, mentre invece, nonostante il valore di *map authentic* sia risultato compatibile con la presenza di un unico profilo anche per il campione C-32, i bassi valori delle frequenze di misincorporazione ottenute, hanno fatto ipotizzare di essere in presenza di un profilo non endogeno, presumibilmente contaminante.

Il programma *ContamMix* ha fornito anche un’elaborazione grafica dei risultati (un esempio per i campioni C-21 e C-35 in Figura 14) generando un grafico a campana il cui picco è dato dalla media dei dati ottenuti e corrisponde al valore di *map authentic* (asse delle ascisse). La base di tale picco risulta essere più o meno slargata a seconda di quanto i dati si discostano dal valore medio.

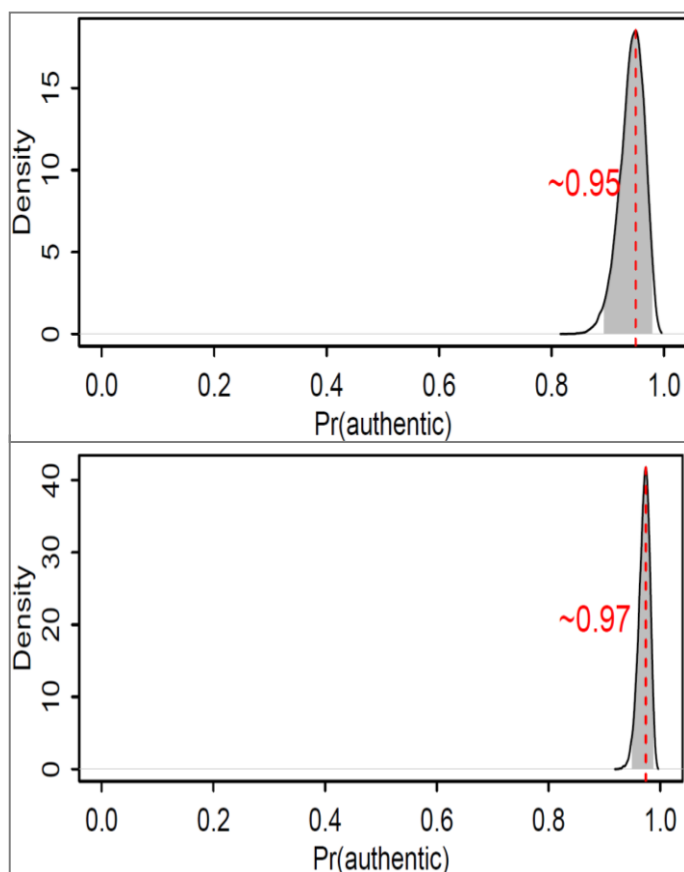


Figura 1: Profilo *ContamMix* relativo ai campioni C-21 e C-35.

Le sequenze consenso di tutti i campioni, tranne del campione C-32 per il quale i valori di *map damage* non erano risultati compatibili con la degradazione del campione, sono state utilizzate per la valutazione dell'aplogruppo mitocondriale mediante il programma *HaploFind*.

HaploFind è in grado di determinare le mutazioni presenti nella sequenza consenso del campione rispetto alla sequenza di riferimento *rCRS* e

associare ad essa l'aplogruppo di appartenenza. La Tabella 12 riporta i risultati ottenuti.

Campioni	Aplogruppo mitocondriale	Score
C-33	N1b1a1	0.70
C-34	U1a1	1.00
C-21	H1h1	1.00
C-23	H	1.00
C-35	T2n	0.50

Tabella12: Aplogruppi mitocondriali dei calci analizzati e *score* assegnato dal programma utilizzato per l'analisi.

Come si può osservare dalla Tabella 12, oltre all'aplogruppo mitocondriale è stato riportato un valore (*score*) che rappresenta un punteggio associato alla probabilità di assegnazione corretta dell'aplogruppo. I dati riportati in Tabella evidenziano come l'assegnazione dell'aplogruppo sia praticamente certa (100%) per i campioni C-34, C-21 e C-23, mentre invece sia inferiore 70% e 50% rispettivamente per i campioni C-33 e C-35.

Come è stato possibile osservare, i campioni che presentavano una copertura della sequenza elevata hanno mostrato anche un valore elevato di *score* mentre il campione C-35 che presentava una copertura inferiore (94.37% all'1 *fold*) ha mostrato un punteggio più basso. Come si può evincere dalla Tabella 12 tutti i campioni presentano un differente

aplogruppo: il campione C-33 appartiene all'aplogruppo N1b1a1, una sottoclasse dell'aplogruppo N1b che identifica le regioni dell'Eurasia (regione compresa tra Europa e Asia) in particolare Russia, Armenia, Tunisia, Turchia, Somalia, Dubai e Ashkenazi Jews; il campione C-34 appartiene invece all'aplogruppo U1a1, una sottoclasse dell'aplogruppo U distribuita tra Nord Europa, Europa Orientale, Asia Centrale, Sud Asiatico, Nord Africa e Corno d'Africa; il campione C-21 appartiene all'aplogruppo H1h1, sottoclasse dell'aplogruppo H, particolarmente frequente nell'Europa Occidentale; il campione C-23 appartiene all'aplogruppo europeo H ed infine il campione SKC-35 appartiene all'aplogruppo T2n, sottoclasse dell'aplogruppo T originario dell'Europa, del Medio Oriente e del Nord Africa, particolarmente rappresentato in Azerbaijan. Dati i risultati ottenuti relativi al *map damage* ed a causa della bassa copertura (57.45% all'1 *fold*), nessun aplogruppo è stato determinato per il campione C-32.

Riassumendo, per quattro campioni (C-33, C-34, C-21 e C-23) su sei è stato possibile ricostruire una sequenza mitocondriale autentica “antica”, per il campione C-32 invece non è stato possibile ottenere informazioni circa la sequenza endogena del campione a causa della scarsa copertura del DNA mitocondriale ottenuta. Pertanto, i dati ottenuti hanno fatto supporre che il materiale genetico analizzato sia di derivazione moderna e quindi contaminante. Nonostante la copertura quasi completa solo all'1*fold* del campione C-35, i valori ottenuti dal *contamMix* e dal *map damage* ci hanno portato ad ipotizzare di essere in presenza di materiale

genetico endogeno “antico”. Inoltre i risultati ottenuti hanno permesso di evidenziare che tutti gli individui analizzati (ad eccezione del campione C-32 di cui non è stato possibile ottenere un risultato) appartenevano ad aplogruppi differenti.

CONCLUSIONI

In questo lavoro preliminare, l'analisi molecolare di campioni ossei altamente degradati provenienti dai calchi rinvenuti nel sito archeologico di Pompei è stata condotta al fine di indagare le eventuali relazioni di parentela per linea materna esistenti fra due differenti gruppi di individui. Una volta puliti e polverizzati i campioni, il DNA è stato estratto e successivamente quantificato. I risultati ottenuti dalla quantificazione hanno permesso di evidenziare che, data la presenza di amplificato nella regione *target* del cromosoma Y, i frammenti ossei analizzati siano presumibilmente appartenuti ad individui di sesso maschile. La sequenza mitocondriale, ottenuta in seguito a cattura del DNA mitocondriale di interesse con sonde oligonucleotidiche specifiche, è stata ricostruita mediante l'utilizzo di *tools* bioinformatici specifici per il DNA antico e successivo mappaggio delle *reads* ottenute con la sequenza di riferimento *rCRS*. I differenti profili mitocondriali ottenuti hanno messo in evidenza la mancata trasmissione, per via materna, del patrimonio genetico fra i vari individui. Dai risultati ottenuti è stato possibile evincere pertanto che i campioni C-34, C-21, C-23 e C-35 non risultano essere imparentati per linea materna. L'impossibilità di ricostruire il profilo mitocondriale del campione C-32 a causa dell'elevato stato di degradazione e della scarsa quantità del materiale genetico estratto non ci ha permesso di identificarne l'aplogruppo mitocondriale di appartenenza e pertanto di stabilire se sia legato da vincoli di parentela, per via materna, al campione C-33.

In conclusione, riassumendo, i risultati genetici che emergono dal questo lavoro hanno messo in evidenza come tutti i campioni analizzati, ad eccezione del campione C-32 per cui non sono stati ottenuti dei risultati attendibili, appartengano ad individui di sesso maschile: è pertanto da escludere l'ipotesi fatta dagli archeologi sull'esistenza di un nucleo familiare caratterizzato dalla presenza di una madre con i suoi figli. Ovviamente con la sola analisi del marcatore mitocondriale a trasmissione matrilineare risulta impossibile valutare l'eventuale presenza nella *Casa del Bracciale d'Oro* di due presunti padri con i loro rispettivi figli oppure di fratellastri per via paterna. Un'ulteriore considerazione deve essere fatta per gli altri due individui analizzati sottolineando che anche una delle due presunte sorelle risulta essere un individuo di sesso maschile e che, dal momento che nessun risultato è stato ottenuto per l'altro campione, nessuna supposizione circa l'eventuale parentela fra i due individui può essere fatta.

BIBLIOGRAFIA

- Alberto Angela** 2016. I tre giorni di Pompei. Rizzoli Vintage.
- Andrews, R.M. Kubacka, I. Chinnery, P.F. Lightowers, R.N. Turnbull, D.M. Howell, N.** 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 23(2):147.
- Briggs, A.W. Stenzel, U. Johnson, P.L.F. Green, R.E. Kelso, J. Prufer, K. Meyer, M. Krause, J. Ronan, M.T. Lachmann, M. Pääbo, S.** 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *PNAS* 104 (37), pp. 14616–14621.
- Cock, P.J. Fields, C.J. Goto, N. Heuer, M.L. and Rice, P.M.** 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, *Nucl. Acids Res.*, 38 (6), pp. 1767-1771.
- Dabney, J. Knapp, M. Glocke, I. Gansauge, M.T. Weihmann, A. Nickel, B. Valdiosera, C. García, N. Pääbo, S. Arsuaga, J.L. Meyer, M.** 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS, USA* 110(39).
- Dabney, J. Meyer, M. and Pääbo, S.** 2013. Ancient DNA Damage. *Cold Spring Harb Perspect Biol.* 5(7): a012567.
- Fu, Q. et al.,** 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Current biology*, 23(7), pp. 553-559.

- Ginolhac, A. Rasmussen, M. Gilbert, M.T. Willerslev, E., Orlando, L.** 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, 27(15), pp. 2153-2155.
- Green, R. et al.** 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*. 134(3), pp. 416-426
- Jonsson, H. Ginolhac, A. Schubert, M. Johnson, P.L.F., Orlando, L.** 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29 (13), pp. 1682–1684.
- Kircher, M.** 2012. Analysis of High-Throughput Ancient DNA Sequencing Data. In: *Ancient DNA: Methods and Protocols, Methods in Molecular Biology.*, ed. Shapiro B. e Hofreiter M., a cura di: Springer Science + Business Media, LLC.
- Li, H. Durbin, R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14), pp.1754-1760.
- Maricic, T. Whitten, M. Paabo, S.** 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE* 5 (11): e14004.
- Meyer, M. Kircher, M.** 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc*; Issue 6.
- Paternò, A. Gatto, F. Marchesi, U., Froiio, A., Verginelli, D., Ruggeri, M.T., Fusco, C., Ciabatti, I., Amaddeo, D.** 2006. Confronto di metodiche real time PCR per l'identificazione e

quantificazione di *Zea mays*. In: Volume degli atti VII Congresso Nazionale S.I.Di.L.V. -Perugia, 9-10-

Sawyer, S. Krause, J. Guschanski, K. Savolainen, V. Pääbo, S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLoS ONE. 7(3)

Vianello, D. Sevini, F. Castellani, G. Lomartire, L. Capri, M. Franceschi, C. 2013. HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum Mutat. 34(9):1189-94.

CONCLUSIONI FINALI

L'attività di ricerca svolta in questi tre anni di dottorato condotta in collaborazione con il Reparto Investigazioni Scientifiche dell'Arma dei Carabinieri di Roma si è concentrata sull'analisi del DNA antico/degradato proveniente, per lo più, da campioni ossei. Tale attività è stata svolta al fine di riuscire ad ottenere un profilo genetico utile per l'identificazione personale sia per poter effettuare una comparazione diretta con un eventuale profilo di confronto, sia per investigare le eventuali relazioni di parentela esistenti fra gruppi di individui antichi.

A causa del maggior numero di copie di DNA mitocondriale all'interno delle cellule del nostro corpo rispetto al DNA nucleare, della struttura circolare che lo rende meno suscettibile all'attacco delle esonucleasi - enzimi idrolitici- e della sua localizzazione all'interno di organelli citoplasmatici dotati di due membrane, il marcatore d'elezione negli studi sul DNA antico è il DNA mitocondriale. L'analisi di quest'ultimo risulta infatti possibile anche quando l'analisi del DNA nucleare fallisce. Tuttavia oggi, grazie allo sviluppo di nuove e maggiormente sensibili tecnologie di analisi è possibile riuscire ad indagare anche il DNA nucleare di campioni degradati. Pertanto, in questo lavoro di tesi, sfruttando le nuove tecnologie di sequenziamento *NGS* ed analizzando il DNA della rocca petrosa, individuato come l'elemento scheletrico "migliore" in termini di qualità e quantità di DNA endogeno, è stato possibile ottenere informazioni genetiche utili ai fini identificativi dall'analisi del DNA nucleare dei campioni antichi/degradati analizzati. Da quanto emerge da questo studio, l'applicazione a campioni

antichi/degradati di nuovi e maggiormente sensibili protocolli forensi quali per esempio quelli per la quantificazione del DNA ha permesso di stimare la quantità del DNA nucleare dei campioni analizzati dando inoltre un'indicazione, ove possibile, sulla degradazione del DNA in essi contenuto e consentendo infine anche la determinazione genetica del sesso. L'utilizzo della tecnologia *NGS* associata a metodi di *target enrichment* differenti (PCR e cattura) ha consentito inoltre, in un caso, di aumentare le informazioni ottenute da un singolo campione, permettendo di effettuare una sorta di *identikit* genetico del campione ignoto e fornendo informazioni sul colore degli occhi, dei capelli e sull'origine geografica; in un altro caso di identificare un profilo genetico utile ai fini identificativi dal materiale biologico digerito prelevato dal tratto gastrointestinale di un insetto ed in un altro caso ancora, di valutare l'autenticità del dato antico sia ottenuto stimando il *pattern* di danneggiamento a cui normalmente va incontro la doppia elica antica che determinando la probabilità che il profilo di DNA ottenuto da ciascun campione appartenga ad un singolo individuo.

Pertanto, in sintesi, da quanto emerge dai risultati espressi nel corpo della presente tesi è stato possibile evidenziare:

1. come le ossa petrose antiche possano essere considerate una buona fonte di DNA ed il DNA estratto da esse possa essere utilizzato per scopi identificativi sfruttando i protocolli impiegati di *routine* in ambito forense. Tale elemento scheletrico potrebbe pertanto essere utilizzato come substrato biologico nei *cold case*;

2. come l'analisi *NGS Ion Torrent* associata alla PCR sia risultata maggiormente sensibile rispetto all'analisi *STR* classica quando applicata a campioni entomologici caratterizzati da DNA degradato e di scarsa quantità. Tale nuova tecnologia ha permesso infatti l'analisi del DNA contenuto all'interno del tratto gastrointestinale di *Pediculus humanus capitis* evidenziando così il supporto che tali risultati genetici possano fornire all'analisi entomologica classica;
3. come l'analisi *NGS Illumina* associata alla PCR abbia permesso, in un'unica reazione, l'analisi di 78 marcatori utili per la determinazione del colore degli occhi e/o l'identificazione dell'origine geografica. Tali informazioni risultano di particolare interesse dal momento che, in assenza di un campione di confronto, possono fornire un "identikit" genetico funzionale all'attività investigativa.
4. come l'analisi *NGS Illumina* associata alla metodica di *target enrichment capture* del DNA mitocondriale di interesse abbia permesso di fornire agli archeologi ulteriori informazioni sugli individui morti a Pompei in seguito all'eruzione vulcanica del 79 d.C. I significativi risultati ottenuti hanno permesso di escludere l'ipotesi fatta dagli archeologici circa l'esistenza di un nucleo familiare imparentato per linea materna fra i calchi in gesso rinvenuti a Pompei.

BIBLIOGRAFIA

- Anderson, S.** 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res*,9 pp.3015-3027.
- Ávila-Arcos, M.C. Cappellini, E. Romero-Navarro, J.A. Wales, N. Moreno-Mayar, J.V. Rasmussen, M. Fordyce, S.L. Montiel, R. Vielle-Calzada, J.P. Willerslev, E.** 2011. Application and comparison of large scale solution-based DNA capture enrichment methods on ancient DNA, *Scientific Reports*,1 (74).
- Briggs, A.W. Good, J.M. Green, R.E. Krause, J. Maricic, T. Stenzel, U. Pääbo, S.** 2009. Primer Extension Capture: Targeted Sequence Retrieval from Heavily Degraded DNA Sources. *Journal of Visualized Experiments*, 31.
- Caramelli, D. Lalueza-Fox, C. Vernesi, C. Lari, M. Casoli, A. Mallegni, F. Chiarelli, B. Dupanloup, I. Bertranpetit, J. Barbujani, G. Bertorelle, G.** 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl Acad. Sci. USA*, 100, pp. 6593–6597.
- Caramelli D., Lalueza Fox C., Condemi S., Longo L., Milani L., Manfredini A, et al.,** 2006. A Highly divergent DNA sequence in a Neandertal individual from Italy. *Curr Biol*. 16, pp. 630-632.
- Cavelier, L. Jazin, E. Jalonen, P. and Gyllensten, U.** 2000. MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. *Hum Genet*. 107: 45–50.

- Cooper, A. Poinar, H.N.** 2000. Ancient DNA: do it right or not at all. *Science*, 289, pp.1139.
- Gilbert, M.T.P. Willerslev, E. Hansen, A.J. Barnes, I. Rudbeck, L. Lynnerup, N. Cooper, A.** 2003 a, Distribution patterns of post-mortem damage in human mitochondrial DNA. *Am. J. Hum. Genet.* 72, pp. 32–47.
- Gilbert, M.T.P. Hansen, A.J. Willerslev, E. Rudbeck, L. Barnes, I. et al.** 2003 b. Characterization of genetic miscoding lesions caused by postmortem damage. *Am. J. Hum. Genet.* 72, pp. 48–61
- Giles, R.E. Blanc, H. Cann, H.M. and Wallace, D.C.** 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77:6715–6719.
- Krause, J. Fu, Q. Good, J.M. Viola, B. Shunkov, M.V. Derevianko, A.P. Pääbo, S.** 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, pp. 894–897.
- Krings, M. Capelli, C. Tschentscher, F. Geisert, H. Meyer, S. et al.** 2000. A view of Neandertal genetic diversity. *Nat. Genet.*, 26, pp. 144–146.
- Lalueza-Fox, C. Sampietro, L.M. Caramelli, D. Puder, Y. Lari, M.** 2005. Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula. *Mol Biol Evol*, 22, pp. 1077–81.
- Lalueza-Fox, C. Krause, J. Caramelli, D. Catalano, G. Milani, L. Sampietro, M.L. Calafell, F. Martínez-Maza, C. Bastir, M. García-Taberner, A. De la Rasilla, M. Fortea, J. Pääbo, S. Bertranpetit, J. Rosas, A.** 2006. Mitochondrial DNA of an Iberian

Neandertal suggests a population affinity with other European Neandertals. *Current Biology*, 16, pp. 629-630.

Orlando, L. Darlu, P. Toussaint, M. Bonjean, D. Otte, M. Hanni, C.

2006. Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr Biol.*, 16, pp. 400-402.

Ovchinnikov, I. Götherström, A. Romanova, G. Kharitonov, V.

Liden, K. Goodwin, W. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature*, 404, pp. 490–493.

Picozzi, M. & Intini, A. 2009. *Scienze Forensi Teoria e prassi dell'investigazione scientifica*. Torino: UTET.

Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak

M, Vernot B, Skov L, Hsieh P, Peyrégne S, Reher D, Hopfe C, Nagel S, Maricic T, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. 358(6363):655-658.

Rasmussen, M. Li, Y. Lindgreen, S. Pedersen, J.S. Albrechtsen, A.

Moltke, I. et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, pp. 757-762.

Rizzi, E. Lari, M. Gigli, E. De Bellis, G. and Caramelli, D. 2013.

Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution* 45:4.

Sanger, F. Coulson, A.R. 1975. A rapid method for determining

sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 94(3):441-8. No abstract available.

-
- Schmitz, R.W. Serre, D. Bonani, G. Feine, S. Hillgruber, F. Krainitzki, H. Pääbo, S. Smith, F.H.** 2002. The Neandertal type-site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl Acad. Sci. USA* 99, pp. 13342–13347
- Skoglund, P. Thompson, J.C. Prendergast, M.E. Mittnik, A. Sirak, K. Hajdinjak, M. Salie, T. Rohland, N. Mallick, S. Peltzer, A. Heinze, A. Olalde, I. Ferry, M. Harney, E. et al.** 2017. Reconstructing Prehistoric African Population Structure. *Cell*. 171(1):59-71.e21.
- Sutovsky, P. Moreno, R.D. Ramalho-Santos, J. Dominko, T. Simerly, C. Schatten, G.** 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod.* 63(2):582-90.
- Tagliabracci, A.** 2010. *Introduzione alla Genetica Forense*. Springer.
- Unterländer, M. Palstra, F. Lazaridis, I. Pilipenko, A. Hofmanová, Z. Groß, M. Sell, C. Blöcher, J. Kirsanow, K. et al.** 2017. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat Commun.* 8:14615.