

**UNIVERSITÀ DEGLI STUDI DI  
FIRENZE**

**DIPARTIMENTO DI BIOLOGIA**

**DOTTORATO DI RICERCA IN ETOLOGIA  
ECOLOGIA ANTROPOLOGIA E  
BIOSISTEMATICA  
(XXVIII CICLO, BIO 08)**

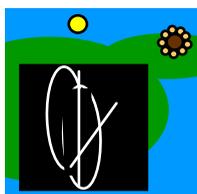
**Caratterizzazione genetica di un osso  
frontale umano arcaico rinvenuto nei  
sedimenti del fiume Po nei pressi di  
Cremona**

**Tesi di**

**Elena Gigli**

**Coordinatore Prof. Alberto Ugolini  
Tutor Prof.ssa Martina Lari**

**(2015)**





*A Gabriele*  
*Alla mia Famiglia*



## INDICE

<b>INTRODUZIONE</b>	<b>pag.4</b>
1.1 Il ritrovamento di Paus	
1.2 Il DNA antico e le sue problematiche	pag 6
1.2.1 Lo studio del DNA antico	pag 6
1.2.2 Caratteristiche del DNA antico	pag 7
1.2.3 Problemi dello studio del DNA antico	pag 10
<b>2 L' ANALISI DEL DNA ANTICO: METODOLOGIE</b>	<b>pag 13</b>
2.1 Next generation technologies	pag 13
2.1.1 Next generation sequencing NGS	pag 13
2.1.2 Strategie di sequenziamento	pag 19
2.2 Criteri di autenticità	pag 21
<b>3 SCOPO DEL LAVORO</b>	<b>pag 26</b>
<b>4 MATERIALI E METODI</b>	<b>pag 27</b>
4.1 Campionamento	pag 27
4.2 Ambiente di lavoro	pag 28
4.3 Pulizia e polverizzazione	pag 28
4.4 Estrazione del Dna da polvere d'osso	pag 29
4.5 Preparazione della libreria a doppio filamento	pag 32
4.5.1 Blunt-end repair	pag 33
4.5.2 Adapter-ligation	pag 33
4.5.3 Adapter fill-in	pag 34
4.5.4 Real-Time PCR	pag 34
4.5.5 Indexing PCR	pag 35
4.5.6 Real time PCR post indexing	pag 37
4.5.7 Accuprime PCR	pag 37
4.6 Preparazione della libreria a singolo filamento	pag 38
4.6.1 Decontaminazione degli adattatori	pag 39
4.6.2 Denaturazione e defosforilazione del primo adattatore	pag 39
4.6.3 Ligazione dei prodotti alle biglie magnetiche	pag 39
4.6.4 Annealing ed estensione dei primers	pag 39
4.6.5 Ligazione del secondo adattatore ed eluizione finale	pag 40
4.6.6 Indicizzazione delle librerie a singolo filamento	pag 40
4.6.7 PCR-1cycle	pag 40
4.7 Cattura del DNA mitocondriale nelle librerie a doppio filamento	pag 41
4.7.1 Preparazione delle sonde	pag 41
4.7.2 Legame delle sonde alle biglie magnetiche	pag 42
4.7.3 Ibridazione fra librerie e sonde	pag 42
4.7.4 Eluizione delle librerie catturate	pag 43

4.7.5 Quantificazione e arricchimento del DNA catturato	pag 44
4.8 Cattura del DNA mitocondriale nelle librerie a singolo filamento	pag 45
4.8.1 Preparazione delle sonde	pag 45
4.8.2 Ibridazione tra libreria e sonde	pag 46
4.8.3 Preparazione delle biglie	pag 46
4.9 Sequenziamento	pag 47
4.10 Analisi bioinformatica	pag 48
4.10.1 Analisi bioinformatica dei risultati del sequenziamento della libreria a doppio filamento	pag 48
4.10.2 Analisi bioinformatica dei risultati del sequenziamento della libreria a singolo filamento	pag 50
<b>5 RISULTATI</b>	<b>pag 52</b>
5.1 Pulizia e polverizzazione	pag 52
5.2. Librerie a doppio filamento	pag 53
5.2.1 Cattura del DNA mitocondriale	pag 56
5.2.2 Analisi bioinformatica	pag 58
5.3 Estrazione e libreria a singolo filamento	pag 63
5.3.1 Shotgun e cattura del DNA mitocondriale	pag 64
5.3.2 Analisi bioinformatica	pag 65
<b>6 DISCUSSIONE</b>	<b>pag 69</b>
<b>7 BIBLIOGRAFIA</b>	<b>pag 71</b>

## RIASSUNTO

Nel 2009 un osso frontale umano emerso dai sedimenti fluviali è stato ritrovato lungo una spiaggia del fiume Po in corrispondenza della confluenza con il fiume Adda. Il reperto, denominato “Paus”, è stato affidato ai curatori del vicino Museo Paleoantropologico del Po. Nonostante la frammentarietà del reperto è apparso subito chiaro che il frontale presentava caratteristiche di arcaicità non attribuibili all’uomo moderno. Nel corso degli anni, nella stessa zona, sono stati ritrovati numerosi resti di fauna tipicamente glaciale che hanno permesso di ipotizzare la presenza di una giacitura primaria di fossili glaciali localizzata molto probabilmente nella anse del fiume Adda. Tuttavia fino alla scoperta di questo osso frontale non c’erano indicazioni di una frequentazione umana della aree padane da parte di ominidi arcaici quali nello specifico i Neandertaliani. Data l’unicità del reperto e la mancanza di contestualizzazione si è reso pertanto necessario avviare un progetto di studio multidisciplinare basato su metodologie all’avanguardia per arrivare a una descrizione approfondita del reperto e a una sua corretta classificazione. In questa tesi verranno descritte le analisi genetiche effettuate su una piccola porzione di materiale osseo al fine di caratterizzare il DNA mitocondriale del reperto. Tale marcatore si è rivelato infatti uno strumento molto utile per classificare le diverse forme umane lungo la linea evolutiva degli ominidi. Per effettuare le analisi paleogenetiche è stato utilizzato un approccio di nuova generazione basato su: i. la preparazione di librerie genomiche; ii. la cattura dell’intero genoma mitocondriale tramite ibridazione con sonde biotilate; iii. il sequenziamento NGS (Next Generation Sequencing); iv. l’analisi delle sequenze ottenute tramite pipeline bioinformatiche specifiche per valutare e massimizzare il livello di autenticità dei dati ottenuti. I dati molecolari ottenuti hanno permesso di dimostrare che, nonostante le pessime condizioni di preservazione e l’elevata contaminazione con DNA esogeno moderno, l’osso di Paus contiene ancora tracce di DNA endogeno e che questo DNA presenta caratteristiche neandertaliane. Questo risultato, oltre a contribuire in maniera sostanziale alla descrizione e classificazione del reperto di Paus, rappresenta la prima prova diretta che questi ominidi frequentassero nelle epoche glaciali anche le aree della pianura padana.

# 1. INTRODUZIONE

## 1.1 Il ritrovamento di Paus

L'uomo di Neandertal si è esteso geograficamente dall'Europa all'Asia in un periodo di tempo compreso tra 130,000 e 30,000 anni BP (Weaver et al. 2009; Hublin 1998). Osservando in un ampio contesto i reperti fossili di questi ominidi risulta evidente che la loro evoluzione sia stata enormemente complessa. Il modello più accreditato per spiegare l'evoluzione di *Homo neanderthalensis* è quello denominato "Accretion Model" (Rosas et al. 2006; Hublin 2009) il quale prevede la comparsa progressiva della morfologia neandertaliana a partire da 450,000 anni BP. In accordo con questa ipotesi la linea neandertaliana è rimasta particolarmente isolata nel territorio europeo inseguito anche allo sviluppo delle condizioni glaciali durante il Quaternario. Questo isolamento geografico ha permesso di fissare progressivamente gli aspetti caratteristici della morfologia neandertaliana attraverso processi di deriva genica e selezione naturale (Hublin 1998).

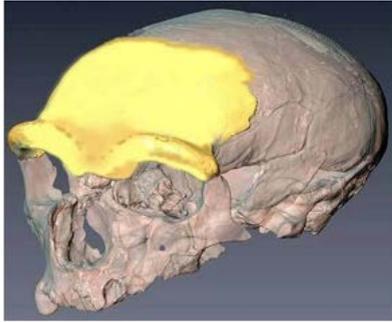
In Italia sono stati rinvenuti numerosi reperti scheletrici di tipo neandertaliano. Fra i più noti ci sono i ritrovamenti del Monte Circeo e di Saccopastore nel Lazio e della Caverna delle Fate in Liguria. Alcuni reperti di dubbia attribuzione sono stati caratterizzati geneticamente. Dall'analisi di alcuni frammenti cranici dei Monti Lessini in Veneto sono emerse importanti informazioni sui tratti fenotipici associati al colore della pelle e dei capelli in questi ominidi (Lalueza-Fox et al. 2006), mentre l'uomo di Altamura in Puglia rappresenta il più antico neanderthal mai analizzato con un'età compresa tra 130,000 e 170,000 anni BP (Lari et al. 2015).

La maggior parte dei siti neandertaliani sono localizzati nel centro e nel sud dell'Italia, mentre per quanto riguarda il nord le uniche attestazioni riguardano i siti liguri e i frammentari reperti dei Monti Lessini, inoltre tutta l'ampia area della pianura padana risulta priva di qualunque testimonianza di frequentazione da parte di questi ominidi. Per questo motivo quando nella primavera del 2009 il giornalista Fulvio Stumpo trovò nei sedimenti del fiume Po nei pressi di Cremona un osso frontale ominoideo (Fig.2) con evidenti caratteristiche di arcaicità, risultò un evento sorprendente.



**Fig.1:** Localizzazione dei reperti neandertaliani in Italia: Veneto: Monte Lessini, Liguria: Caverna delle Fate. Toscana: Buca del Tasso. Lazio: Grotta Guattari, Grotta Breuil, Grotta del Fossellone. Campania: Roccia San Sebastiano. Puglia: Altamura, Grotta Santa Croce. Calabria: Ianni di San Calogero, San Francesco di Archi.

Il reperto, a cui è stato dato il nome di “Paus”, è stato trovato in una barra fluviale del Po, in particolare in una spiaggia di fronte alla confluenza del fiume Adda poco a monte di Cremona. Purtroppo si trovava in una posizione secondaria senza quindi riferimenti stratigrafici e in assenza di manufatti o resti che potessero aiutare nella contestualizzazione. Nel corso degli anni, nella stessa zona, sono stati ritrovati numerosi resti di fauna tipicamente glaciale come cervo megacero, bisonte, mammut ed alce, abbondanti presso l’abitato di Spinadesco nel fiume Po e sulle barre di meandro a sud di Pizzighettone (CR) nell’alveo dell’Adda. Recenti studi bibliografici e sul territorio stesso hanno permesso di ipotizzare la presenza di una giacitura primaria di fossili glaciali formata da uno spessore variabile di strati torbosi alternati a strati di argille azzurre compatte e presente in due anse del fiume Adda. Risulta quindi possibile un passaggio di fossili che dal fiume Adda, dopo trasporto, si andrebbero a depositare sulle prime barre del Po, dopo confluenza. Il reperto di Paus si va quindi a collocare in un quadro compatibile con la presenza dell’uomo di Neanderthal nella Pianura Padana, ma in seguito alle particolari condizioni di ritrovamento, alla mancanza di altri reperti o manufatti associati e di informazioni relative al sito di deposizione primaria, sono necessarie delle analisi approfondite sia in ambito morfometrico che genetico per una corretta classificazione e descrizione del reperto.



**Fig.2:** In giallo l'osso frontale appartenente a Paus

## **1.2 Il DNA antico e le sue problematiche**

### **1.2.1 Lo studio del DNA antico**

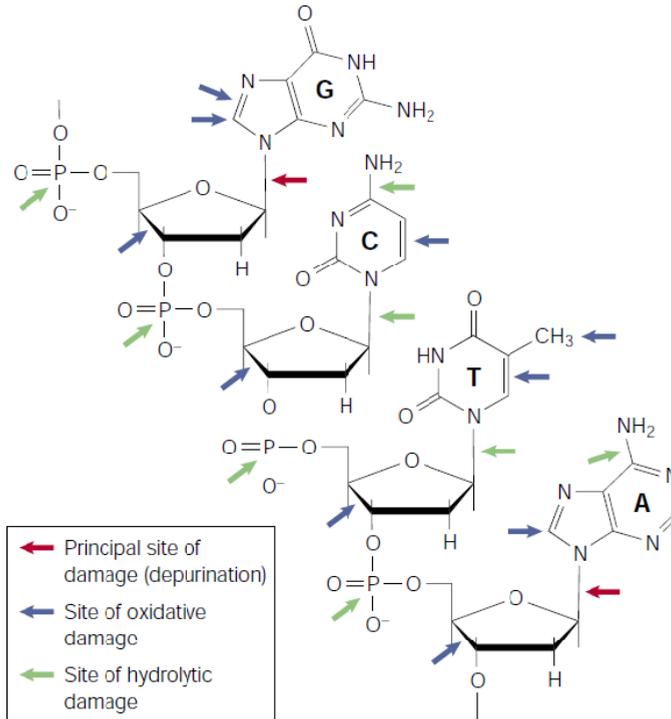
L'analisi del DNA antico è un campo di studio con elevatissima possibilità di applicazione che si è sviluppato negli ultimi anni soprattutto grazie al progresso continuo delle tecniche di biologia molecolare. Si parla di DNA antico (*ancient DNA* o aDNA) quando si ha a che fare con qualsiasi residuo di materiale genetico che può essere estratto da una grande varietà di reperti, diversi per tipo, età, origine e stato di conservazione (Rollo et al. 1999), senza limitarsi a quello estratto da reperti antichi ma comprendendo un qualunque acido nucleico con struttura alterata o degradata. Nel DNA non solo sono scritte tutte le informazioni genetiche che influenzano i caratteri fenotipici di un organismo, dall'embrione fino alla capacità di adattarsi all'ambiente esterno, nella sua sequenza resta leggibile anche la sua storia naturale. Per questo poter studiare il DNA antico recuperato da reperti fossili come ossa, tessuti mummificati, coproliti (resti fecali antichi), semi di piante e così via, può permettere la comprensione di eventi avvenuti anche milioni di anni fa. Attraverso lo studio di specifici marcatori è possibile identificare e confrontare sequenze genetiche significative di diverse specie viventi e rapportandole con quelle moderne, studiare come le caratteristiche genetiche cambiano nel tempo, come erano e sono diverse da un luogo ad un altro fino a ricostruire le origini e i flussi migratori di intere popolazioni. Ovviamente il presupposto è che il DNA si conservi in questi reperti attraverso i secoli e i millenni. Quale sia il limite della vita del DNA è stato a lungo argomento di discussione fra gli scienziati. Varie pubblicazioni, riguardanti sequenze antiche ottenute da resti di dinosauri (Woodward et al. 1994), da una pianta di 17 milioni di anni fa (Golenberg et al. 1990) e da insetti inglobati in ambra risalenti a circa 120-135 milioni di anni fa, sono state rapidamente confutate e ritenute non veritiere per la presenza di contaminazioni o del fallimento di repliche indipendenti (Zischler et al. 1995; Austin et al. 1997; Walden e Robertson 1997), fino alla consapevolezza che il DNA non è eterno ma come ogni organismo invecchia e muore. Alcuni scienziati hanno

imposto come limite temporale alla conservazione del DNA nei campioni fossili 100.000 anni (Pääbo et al. 2004) più precisamente è la temperatura della zona di conservazione che modula questo limite, frammenti lunghi 100-500 paia di basi (bp) possono sopravvivere più di 10.000 anni in regioni temperate e massimo 100.000 anni a climi più freddi (Pääbo e Wilson 1991; Poinar et al. 1996; Smith, Chamberlain et al.2001). Questo, in generale, è il limite accettato dalla maggior parte degli studiosi anche se esistono lavori in cui è stato ottenuto DNA amplificabile da campioni recuperati da sedimenti ghiacciati datati 300.000-400.000 anni fa (Willerslev et al. 2003). Il primo studio convincente effettuato sul DNA antico risale al 1984, un'équipe americana diretta da Russel Higuchi ha estratto e sequenziato materiale genetico da un frammento di muscolo essiccato di *Equus quagga quagga* un equide simile alla zebra estinto nel XIX secolo. Un anno più tardi è lo svedese Svante Pääbo a fornire la prima dimostrazione che lo studio dell' aDNA è possibile anche in resti umani, analizzando sequenze di DNA appartenenti a un' antica mummia egiziana (Pääbo et al. 1985). Entrambi i lavori effettuati con metodologia classica, hanno dimostrato che il materiale estratto è per la maggior parte di origine fungina e microbica. Generalmente, infatti, il DNA estratto da campioni antichi si presenta in poche copie di frammenti brevi e molto danneggiati (Dabney et al. 2013). Con l' avvento della *Polymerase Chain Reaction* (Reazione a catena della polimerasi, PCR), lo studio del DNA antico ha ricevuto una notevole spinta in avanti, questa tecnica rende possibile un' amplificazione selettiva del materiale genetico, teoricamente anche a partire da una singola molecola di DNA (Pääbo et al. 1989). Oltre al basso numero di copie in cui si trova, il DNA antico presenta molte altre caratteristiche che complicano il suo studio e con le quali mette alla prova l' antropologo molecolare.

### **1.2.2 Caratteristiche del DNA antico**

Alla morte di un organismo vengono meno tutti i processi di riparazione, tipici di una cellula vivente, necessari a preservare l' integrità del materiale genetico, da questo momento la molecola di DNA diventa instabile e va incontro a tutta una serie di processi degradativi definiti complessivamente "diagenesi" che ne alterano la struttura e le proprietà chimico-fisiche e che nel tempo portano alla sua completa degradazione. Il DNA estratto da reperti antichi appare degradato in frammenti di piccole dimensioni, stimate tra 40 e 160 bp con le metodiche più recenti (Sawyer et al. 2012) anche se in studi precedenti sono stati recuperati frammenti fino a 500 bp (Pääbo 1989). Il processo di frammentazione è dovuto sia a reazioni enzimatiche, che avvengono subito dopo la morte dell' organismo come l' azione delle nucleasi lisosomiali, l' attacco di batteri, funghi e insetti che si nutrono dei tessuti, sia a reazioni non enzimatiche, influenzate da vari fattori che portano a una degradazione spontanea della doppia elica. La principale reazione di degradazione non enzimatica è l'idrolisi dei legami N-glicosidici tra la base

azotata e lo zucchero 2-deossiribosio, con conseguente perdita della base e creazione di un sito abasico. Poichè a condizioni fisiologiche questo processo avviene con una velocità 20 volte superiore per adenina e guanina, questa reazione prende il nome di depurinazione. In corrispondenza di un sito apurinico il doppio filamento è indebolito e può avvenire un riarrangiamento chimico che si risolve nella rottura della doppia elica e quindi nella frammentazione della molecola di DNA. L'idrolisi può colpire anche il legame fosfodiesterico dello scheletro zucchero fosfato, causando la formazione di un *nik* a singolo filamento nella doppia elica che la destabilizza e può portare alla sua rottura (Lindhal et al. 1993). L'idrolisi può colpire anche i gruppi amminici delle basi stesse: le basi azotate con gruppi amminici secondari la citosina, il suo omologo 5-metil citosina e la guanina vanno incontro a deaminazione diventando rispettivamente uracile, timina e ipoxantina (Hofreiter et al. 2001; Poinar 2002). La deaminazione è ritenuta la principale causa di *lesioni miscoding* nel DNA antico (Briggs et al., 2007) essa infatti provoca l'incorporazione di basi diverse durante il processo di amplificazione (Pääbo 1989; Lindha, 1993; Willerslev e Cooper 2005). Anche l'ossigeno interviene nella diagenesi del DNA, ed è in grado di causarne la frammentazione. Esso è infatti responsabile della produzione di radicali liberi come i radicali di perossido ( $O_2$ ), perossido di idrogeno ( $H_2O_2$ ) e i radicali idrossilici (OH) (Lindhal 1993), in grado di convertire la guanina in 8-idrossi-guanina e di dare origine a derivati saturi delle pirimidine, forme idrate e glicoli. Questa ossidazione delle pirimidine può portare alla rottura della doppia elica. I radicali possono essere prodotti dalle cellule stesse e da radiazioni ionizzanti oppure risultare da processi cellulari degradativi avviati da funghi e batteri (Poinar 2002). Oltre all'ossidazione, le pirimidine esposte a radiazioni ultraviolette formano legami coniugati tra loro stesse creando dimeri di pirimidine.



**Figura 3:** I principali danni che colpiscono l'aDNA. Le frecce verdi indicano i siti suscettibili al danno idrolitico; le frecce blu quelli suscettibili al danno ossidativo; le frecce rosse la depurinazione.

Un altro tipo di danno subito dal DNA è rappresentato dalla formazione di cross-link intra e intermolecolari, un esempio sono i prodotti di Mallard, originati dalla condensazione fra zuccheri e gruppi amminici primari di proteine e acidi nucleici, presenti soprattutto nei coproliti. Giocano un ruolo importante nei danni al DNA anche pH, temperatura e forza ionica del mezzo di conservazione del reperto. La variazione di questi fattori influenza le reazioni di idrolisi e di ossidazione aumentandone o diminuendone l'effetto sulla sequenza nucleotidica, pertanto lo stato di degradazione del DNA e l'estensione della frammentazione risultano più o meno spinti a seconda delle condizioni ambientali in cui si è conservato il DNA. Un pH acido aumenta la deaminazione (Mitchel et al. 2005), mentre un pH basico aumenta la possibilità di rottura del filamento e quindi la creazione di siti abasici (Lindhal 1993); le basse temperature inibiscono l'attività enzimatica e preservano meglio l'integrità del DNA, come dimostrato dallo stato di conservazione di vari campioni ritrovati a climi molto freddi tra cui: la mummia del Similaun, nota anche come Uomo del Similaun (uomo venuto dal ghiaccio e informalmente Ötzi) datata 3330 BP, è stata ritrovata in un ghiacciaio sulle Alpi Venoste, (Handt et al. 1994; Rollo et al. 1994, 1995,

1999), un esemplare di mammut (Hoss et al., 1994; Gilbert et al. 2007; Poinar et al. 2006), di bisonte (Gilbert et al. 2004; Shapiro et al. 2004) ed un cloroplasto (cpDNA) ritrovati nel permafrost. Le alte temperature, invece, indeboliscono le forze di Van der Waals e i legami a idrogeno, oltre a promuovere la crescita di batteri funghi e altri microorganismi che contribuiscono alla degradazione e alla contaminazione del DNA endogeno del campione (Handt et al. 1994; Rollo 1999; Caramelli e Lari 2004; Gilbert et al. 2007).

### **1.2.3 Problemi nello studio del DNA antico**

#### **a) Frammentazione**

Il danno *post-mortem* più evidente a cui l'antropologo molecolare si trova di fronte, è la frammentazione, dovuta all'attivazione di endonucleasi cellulari o conseguenza di danno idrolitico o ossidativo. I nuovi metodi di sequenziamento (Margulies et al. 2005) e il nuovo sistema di cattura (Briggs et al. 2009; Burbano et al. 2010; Maricic et al. 2010) hanno dimostrato che più della metà dei frammenti di DNA estratti dai resti antichi sono di dimensione inferiore ai 50 nucleotidi e quindi non recuperabili attraverso la PCR.

#### **b) Contaminazione**

Un'altra delle problematiche da affrontare nello studio dell'aDNA è la contaminazione da DNA esogeno, è la "minaccia" principale soprattutto quando si tratta di campioni antichi umani. La capacità della PCR, di rilevare in modo specifico anche piccole quantità di DNA e amplificarle milioni di volte in pochissime ore, ha dato un notevole impulso allo studio del DNA antico, ma si rivela un'arma a doppio taglio nella contaminazione. Essa, con la sua ipersensibilità, è capace di amplificare anche DNA esogeno contaminante e generare falsi positivi, soprattutto se, come già accennato, il DNA endogeno d'interesse è solo una minima percentuale dell'estratto totale. Le fonti di contaminazione variano a seconda del tipo di campione antico in esame. Il DNA antico animale e vegetale è spesso contaminato da campioni moderni usati come confronto durante le identificazioni morfologiche dei resti (Yang e Watt Kathy 2005). Il DNA di batteri patogeni antichi può essere contaminato da quello di specie strettamente correlate presenti in terreni e ambienti circostanti (Greenblatt e Spigelman 2003). Nel caso del DNA antico umano, è il DNA moderno la principale fonte di contaminazione. Il rischio di contaminazione da DNA moderno è praticamente costante durante le varie fasi che si susseguono nello studio di reperti antichi. Una manipolazione non attenta dei campioni può causare contaminazione del campione con cellule epiteliali dell'operatore (Gilbert et al. 2005a; Fortea et al. 2008). Nelle fasi recupero e durante le analisi morfometriche le principali fonti di possibile DNA contaminante sono l'archeologo e l'antropologo fisico che manipolano il reperto (Sampietro et al. 2006), successivamente anche l'antropologo molecolare che si occupa di

decontaminarlo con raggi UV, prima di procedere con l' estrazione. Qualsiasi operatore che manipola il reperto dovrebbe indossare guanti, camice e mascherina sterili per evitare la contaminazione con DNA esogeno (Cooper e Poinar 2000; Caramelli e Lari 2004). Durante la procedura di estrazione, la fonte principale di contaminazione è l'operatore. Il DNA contaminante può essere introdotto nella preparazione dei reagenti, ma può anche derivare da contaminazione ambientale del luogo in cui si esegue l'operazione. Per evitarla l'ambiente di lavoro deve essere sottoposto ad accurati accorgimenti, luce UV accesa durante la notte e pulizia continua dei piani di lavoro con candeggina o alcool. La preparazione di tutti i reagenti deve essere fatta sotto una cappa e gli operatori devono obbligatoriamente utilizzare tuta, guanti, mascherine e cuffie sterili (come già precedentemente detto). Durante la fase di amplificazione (PCR) il DNA contaminante può essere introdotto principalmente in due momenti: durante l'allestimento della reazione e durante la reazione vera e propria. Nell'allestimento della reazione infatti si possono contaminare i reagenti oppure errori manuali dell'operatore possono portare alla contaminazione della mix di reazione. Per questo anche in questa fase è obbligatorio usare le stesse norme utilizzate nella fase di estrazione. Nella reazione invece, spesso vengono usate condizioni aspecifiche come bassa temperatura di annealing e elevato numero di cicli per cercare di ovviare alla bassa quantità di aDNA presente. Questo può causare appaiamento dei primer con sequenze esogene presenti nell'estratto e loro amplificazione al posto del DNA d' interesse.

### c) Inibizione

Il fenomeno dell' inibizione, può presentarsi in fase di amplificazione ed è legato proprio alla grande sensibilità della PCR. Diversi fattori (detti inibitori) sono infatti in grado di bloccare l'azione della *Taq polimerasi* impedendo così la reazione di amplificazione (Pääbo 1989; Hanni 1995; Poinar 2002). Gli inibitori provengono da una fonte esterna (Scholz et al. 1998; Poinar 2002) e possono essere:

- Alcuni danni ossidativi che possono bloccare o alterare la progressione della *Taq polimerasi* (Mitchell et al. 2005; Scholz et al. 1998)
- Il DNA non bersaglio principalmente di origine batterica (Pääbo 1989)
- Ioni complessati come ferro (Fe) o componenti del suolo come tannini e acidi umici (Pääbo 1989; Hagelberg et al. 1991; Poinar 2002; Scholz et al. 1998)

Per superare questo problema sono utilizzati diversi metodi come ad esempio l'aggiunta di albumina di siero bovina (*Bovine Serum Albumine*, BSA) alla mix di reazione che cattura gli inibitori e indirettamente promuove l'attività della polimerasi oppure durante l'estrazione l'aggiunta del DNA N-phenacyltiazoliumbromide (PTB) che favorisce la rottura dei prodotti di Maillard, liberando il DNA per l'amplificazione (Poinar 2002). Oppure i

campioni possono essere diluiti con H<sub>2</sub>O distillata così da ridurre la concentrazione degli inibitori.

#### d) Lesioni di miscoding

Oltre a quanto detto, l'analisi corretta della sequenza può essere complicata dalle *lesioni di miscoding*, un'errata incorporazione delle basi causata dai prodotti delle reazioni di idrolisi che portano a depurinazione e deaminazione. Come già visto è la deaminazione la causa principale di queste lesioni. Le *lesioni di miscoding* sono state analizzate in vari studi (Pääbo 1989; Hofreiter et al. 2001; Gilbert et al. 2003a; 2003b), molti dei quali hanno mostrato che la maggior parte dei cambiamenti derivano dalla deaminazione della citosina (C) in uracile (U), analogo della timina (T), o dalla deaminazione di adenina (A) in ipoxantina (HX), analogo della guanina (G). Ognuna delle due transizioni è in grado di produrre due fenotipi osservabili sulla sequenza a seconda di quale dei due filamenti complementari viene sequenziato. La deaminazione di C in U può essere osservata come una semplice transizione C→T, oppure sul filamento complementare come G→A. Analogamente la deaminazione di A in G può essere letta come A→G o come T→C sull'altro filamento. Con il lavoro di Hansen e colleghi (Hansen et al. 2001) questi set di lesioni sono state divisi in transizioni di "tipo 1" (A→G / T→C) o di "tipo 2" (C→T / G→A). Le nuove tecnologie di sequenziamento hanno permesso di indagare più a fondo la natura di questi cambiamenti (Gilbert et al. 2007) mostrando come siano prevalentemente le "lesioni di tipo 2" quelle rappresentative dei danni al DNA. Il lavoro di Brotherton e colleghi (Brotherton et al. 2007), sfruttando la messa a punto di una nuova tecnica Single Primer Extension (SPECT), ha fornito una forte evidenza quantitativa mostrando che solo la modifica C→U/T è la causa di un danno endogeno al DNA derivante da *lesioni di miscoding*. Mentre le lesioni di "tipo 1" sono quelle non endogene, generate dalla PCR e spariscono usando tecniche alternative come ad esempio la digestione con uracile-N-glicosilasi (UDG) (Stiller et al. 2006; Gilbert et al. 2007; Brotherton et al. 2007). Altra osservazione importante è la diversa distribuzione delle *lesioni di miscoding* all'interno della molecola di DNA. Lo studio di Briggs e colleghi (Briggs et al. 2007) ha mostrato infatti come le deaminazioni della citosina (C→T/ G→A), siano maggiormente concentrate alle estremità dei filamenti (al 5' C→T e al 3' G→A) rispetto alla parte centrale, probabilmente perché le estremità a singolo filamento originate dalla degradazione risultano più sensibili rispetto al doppio filamento. La quantità di queste lesioni varia da reperto a reperto e sono maggiormente presenti, nei reperti dove i danni da idrolisi hanno pesantemente agito sulla struttura del DNA. Tale fenomeno è quindi correlato con l'ambiente in cui il reperto si è conservato.

## 2. L'ANALISI DEL DNA ANTICO: METODOLOGIE

### 2.1 Next generation technologies

#### 2.1.1 Next generation sequencing (NGS)

Le più recenti ed innovative tecnologie biomolecolari e di sequenziamento, denominate Next Generation Sequencing (NGS) sono in grado di aumentare il numero di basi sequenziate per ogni corsa riducendo i costi e generando una quantità di dati di sequenza dell'ordine di Gigabasi in poche ore. Queste caratteristiche le rendono molto più adatte all'analisi di genomi completi o geni molto lunghi e il loro progresso sta drasticamente cambiando il volto della ricerca in diverse aree della biologia. La loro efficienza e versatilità le rende adatte anche per l'analisi del DNA antico, in questo campo hanno quasi completamente sostituito il metodo di Sanger. Questo metodo di sequenziamento *high-throughput* permette la caratterizzazione completa di un estratto di DNA altamente degradato limitando tutte le problematiche relative alla frammentazione, degradazione e basso numero di copie, offrendo così un vantaggio enorme rispetto alla PCR della metodologia classica. Nello studio del DNA antico le piattaforme NGS maggiormente usate sono state, inizialmente la 454/Roche FLX e successivamente ed attualmente la Illumina Genome Analyser. La prima, descritta nel 2005 (Margulies et al. 2005) è stata quasi subito utilizzata per l'analisi dell'aDNA, tuttavia a causa delle procedure più complesse e dei costi più elevati a parità di output di sequenziamento, è stata rapidamente sostituita in questo settore da Illumina

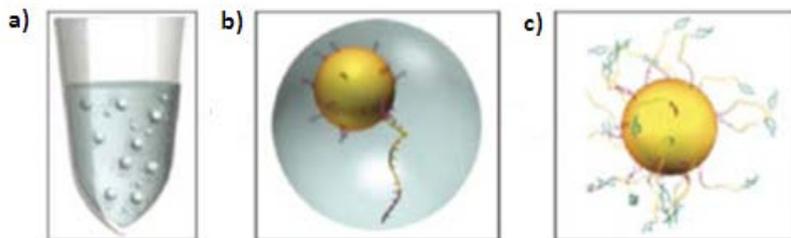
Le due piattaforme di sequenziamento differiscono nella fase di amplificazione e nella tipologia di sequenziamento effettuato, presentando così un flusso di lavoro leggermente diverso. Nonostante l'elevata sensibilità e l'enorme numero di dati che sono in grado di generare, il loro sistema di rilevamento, una camera a CCD, non è capace di leggere il segnale relativo ad una singola molecola di DNA ma, lo individua solo se è generato da milioni di molecole. Per questo nel flusso di lavoro la costruzione di librerie genomiche di DNA target e la loro amplificazione risultano fondamentali.

Sono tre i passaggi chiave caratteristici del *work flow* NGS:

1. **Preparazione delle librerie genomiche:** questa fase consiste nell'applicazione alle estremità dei frammenti di DNA estratti, di due adattatori universali brevi oligonucleotidi con sequenza nota. In realtà il legame degli adattatori è preceduto dalla frammentazione dell'estratto per sonicazione o digestione con enzimi di restrizione ma, dato che il DNA antico è già frammentato questa fase non viene effettuata. Questa stessa caratteristica dell'aDNA rende però necessaria la riparazione di eventuali estremità sfalsate al 5'-3' per generare estremità piatte. Si usano simultaneamente una DNA polimerasi e una chinasi, la prima rimuove o ricostruisce la sequenza

complementare, la seconda catalizza l'aggiunta di gruppi fosfato al 5'. Una volta riparati, ai frammenti sono legati gli adattatori, sui quali sono disegnati i primers necessari alle successive fasi di amplificazione e al sequenziamento.

2. **Amplificazione della library:** in questa fase, i frammenti vengono amplificati utilizzando primers complementari agli adattatori ma le due piattaforme utilizzano tecniche diverse. La differenza principale risiede nel fatto che nella 454 ogni singolo frammento di libreria è amplificato indipendentemente mentre per l'Illumina si producono *cluster* di frammenti amplificati. Nella 454 l'amplificazione avviene per PCR emulsioclonale o emPCR. Ogni molecola di library è legata tramite gli adattatori ad una biglia che fa da supporto per la PCR e la reazione di amplificazione si compie, per ogni frammento, in una gocciolina d'acqua (microreattore o "microcella") dispersa in un solvente oleoso. In questo modo, ogni frammento è isolato dagli altri e si replica in modo indipendente. Alla fine dell'emPCR, l'emulsione viene rotta, gli amplificati denaturati e le biglie disposte nelle celle di una piastra di fibre ottiche, denominata *Pico Titer Plate*, il supporto per il sequenziamento.



**Figura 4:** amplificazione library mediante em-PCR. **a)** biglie magnetiche disperse nell'emulsione. **b)** microcella in cui avviene la reazione e biglia con frammento di DNA legato mediante l'adattatore. **c)** library amplificata

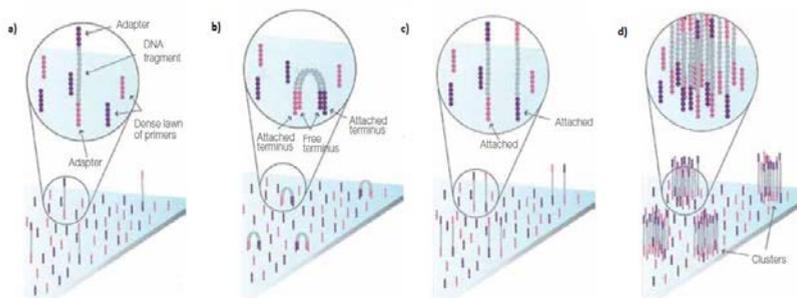
Nell'Illumina invece, l'amplificazione avviene su un supporto solido, detto *flow cell* (Fig.5) un vetrino funzionalizzato con oligonucleotidi complementari agli adattatori.



*flow cell*

**Figura 5:** flow cell, vetrino con otto canali la cui superficie è derivatizzata con oligonucleotidi complementari agli adattatori

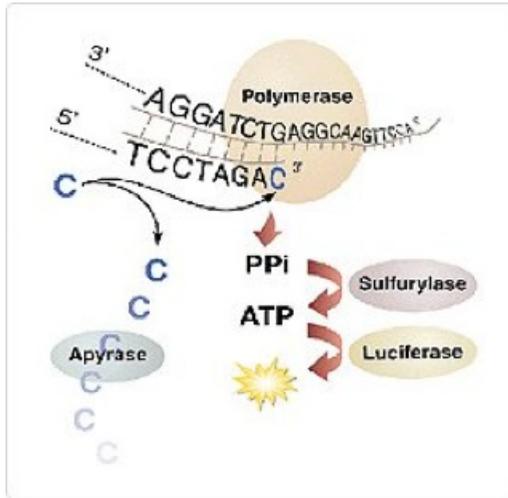
Le molecole di libreria si legano al vetrino e l'amplificazione avviene attraverso una PCR detta *Bridge* (Fig.6) che prevede due meccanismi distinti. Prima gli oligonucleotidi ancorati funzionano da primer, mentre le molecole di libreria a loro ibridizzate da template, per una prima generazione di copie dei filamenti di DNA che restano così ancorate alla superficie; poi dopo la denaturazione i filamenti stampo tornano in soluzione rendendosi di nuovo disponibili per ulteriori processi di ibridazione e allungamento dei primer ancorati. Le stesse copie così generate dei filamenti di libreria, possono ora funzionare, mediante ibridazione "a ponte" con gli oligonucleotidi vicini, da template per un'ulteriore amplificazione, generando copie aggiuntive in prossimità di quelle iniziali (Adessi et al. 2000). Così si generano gruppi di frammenti amplificati riuniti in clusters. Data la densità degli oligonucleotidi presenti sulla superficie del *flow cell*, è possibile ottenere molte copie dello stesso frammento in un'area piccola, si stimano circa mille copie per micrometro di superficie.



**Figura 6:** amplificazione della libreria illumina. **a)** legame dei frammenti alla superficie del flow cell. **b)** PCR-Bridge. **c)** denaturazione dei bridge, così le copie generate, possono funzionare da template per ulteriori amplificazioni. **d)** clusters generati sul flow cell.

A differenza della 454, Illumina mantiene il *flow cell* anche come supporto per il sequenziamento. Per entrambe le tecniche la quantificazione della libreria è un passaggio critico e necessario per ottenere un'alta qualità di sequenza. Mentre per l'Illumina il numero appropriato di clusters è garantito dall'amplificazione, nella 454 è necessaria una precisa quantificazione della library in modo da trovare il giusto rapporto DNA/biglie, tale che solo 1 frammento di DNA si leghi per ogni biglia aumentando il rendimento dell'emPCR.

3. **Sequenziamento:** Anche in questa fase le due tecniche si distinguono seguendo diverse strategie e questo comporta un output diverso in basi sequenziate per corsa. La piattaforma 454, utilizza il *pirosequenziamento*, una metodica basata sul dosaggio del pirofosfato (Ppi) liberato in seguito all'attacco di un dNTP al filamento polimerizzato, la reazione di sequenza avviene in ogni cella della *Pico Titer Plate*. Questo metodo utilizza tre enzimi ATPsolforilasi, luciferasi e apirasi e i substrati adenosin solfo solfato (ASP) e luciferina, con i quali è incubata la doppia elica denaturata. I quattro dNTPs sono aggiunti separatamente per evitare sovrapposizione. Ogni volta che la DNA polimerasi aggiunge il nucleotide complementare, induce rilascio di Ppi che avvia a una serie di reazioni a cascata: la solforilasi, usando l'ASP, trasforma il Ppi in ATP che a sua volta viene usato dalla luciferasi per convertire la luciferina in ossiluciferina con produzione di un segnale luminoso (Fig.7). L'enzima apirasi degrada il dNTP non incorporato e l'ATP prodotto, al termine della degradazione è aggiunto il secondo dNTP e così via fino alla sequenza completa. Gli spot di luce emessi vengono registrati da una camera CCD. L'output è un pirogramma, caratterizzato da picchi di altezza proporzionale al numero di nucleotidi aggiunti e solo successivamente viene convertito in sequenza nucleotidica. Dopo la versione GS20, che detiene diversi successi nello studio del DNA antico (Gilbert et al. 2007), nel 2005 la Roche ha proposto una nuova versione della tecnologia 454, la GS FLX Titanium Plus, capace di generare fino a 1 Gigabase di dati per corsa e di leggere sequenze fino a 800bp.

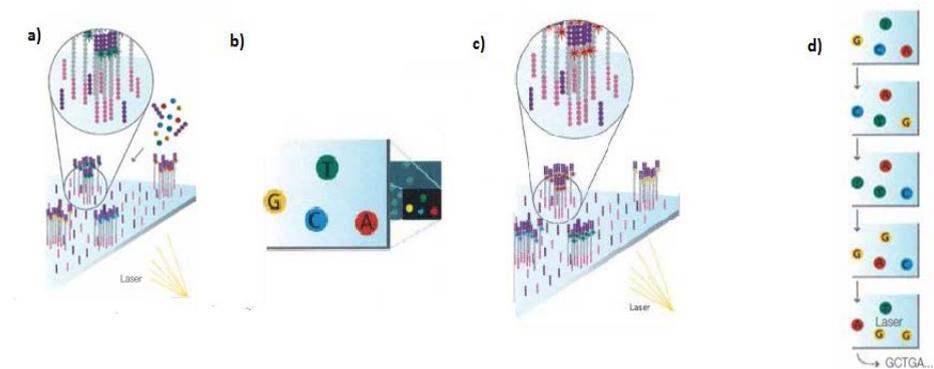


**Figura 7:** Pirosequenziamento della 454-FLX

L'Illumina invece utilizza un sistema di sequenziamento mediante sintesi (SBP, *Sequencing by Synthesis*), nel quale tutti e quattro i nucleotidi sono contemporaneamente applicati ai canali del *flow cell* insieme ad una DNA polimerasi modificata. Più precisamente i nucleotidi sono CTR, *cyclic reversible terminators*, ognuno è marcato con un fluorocromo scindibile specifico per ciascuna base e presenta nel gruppo ossidrilico al 3' un blocco chimicamente eliminabile, così che si verifichi un unico evento di incorporazione. Durante la reazione, i CTR sono eccitati da un laser e ogni evento di incorporazione è seguito da emissione di fluorescenza registrata da una CCD camera, mentre un computer monitora ogni cluster a cui viene assegnata una sequenza. La rimozione del blocco al 3' OH e del fluoroforo prepara ogni filamento all'estensione a singolo nucleotide successiva per rilevare la base seguente. I cicli di incorporazione ed emissione di fluorescenza sono impostati dall'operatore e consentono di ottenere letture di sequenza di lunghezza che varia da 36 a 300 bp a seconda del tipo di analizzatore. L'Illumina GAIIIX *platform* genera circa 50 Gigabasi per corsa, mentre i modelli più recenti molte di più. Si possono ottenere fino a 1000 Gigabasi per corsa con una lunghezza di lettura fino alle 150 bp da entrambe le estremità per la piattaforma HiSeq e fino alle 300 bp da entrambe le estremità nel caso della piattaforma Miseq. Ad oggi per l'analisi del DNA antico è questa la tecnologia più usata, rispetto alla piattaforma 454/Roche, Illumina permette il recupero di dati di sequenza relativi anche a frammenti molto corti e di generare milioni di letture ad un costo minore per base.

Un altro vantaggio delle tecnologie di nuova generazione risiede nella capacità di *multiplexing* grazie alla quale è possibile sequenziare più campioni nella stessa corsa. Oltre a sfruttare le partizioni fisiche delle plate di sequenziamento, si possono infatti creare anche librerie "indicizzate",

legando agli adattatori degli oligonucleotidi di piccole dimensioni (6-10bp), con sequenza nota, diversi per ogni campione. Questi oligonucleotidi sono detti MID per la 454 e indici per l'Illumina. In questo modo diversi campioni possono essere riuniti in unico *pool* (insieme di molti campioni) e sequenziati insieme (in *multiplex*) in un'unica lane. Successivamente, una volta ottenute le sequenze, i MID e gli indici permettono di distinguerli (Fig.8).



**Figura 8:** Sequenziamento di Illumina. **a)** Aggiunta dei quattro CTR in contemporanea insieme a primers complementari agli adattatori e una DNA polimerasi modificata. **b)** determinazione della prima base. **d)** nuova estensione a singolo nucleotide. **c)** alla fine dei cicli di sequenziamento determinazione della sequenza completa.

### 2.1.2 Strategie di sequenziamento

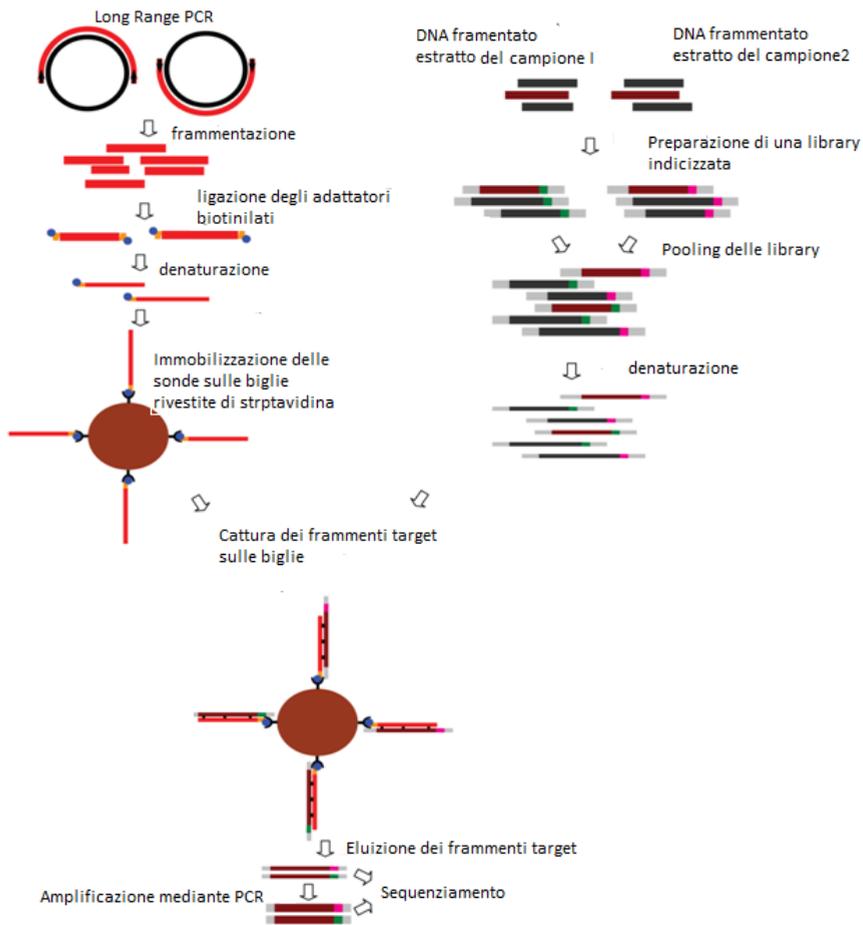
Quando si esegue un sequenziamento *high-throughput* è possibile seguire due strategie: il **metodo Shotgun**, che prevede il sequenziamento dell'intero genoma, ed **metodo Capture** con cui si sequenziano regioni specifiche, "catturate" dal DNA totale attraverso sonde oligonucleotidiche complementari alle regioni d'interesse. Il metodo Shotgun è stato il primo metodo impiegato nell'analisi del DNA antico e offre un ampio spettro di tutto il DNA che costituisce il reperto ma, come già spiegato, la maggior parte dell'aDNA estratto dai reperti è di natura microbica e fungina e questo rende l'approccio inadatto se non a volte decisamente inapplicabile (Burbano et al. 2010). Uno dei maggiori problemi nell'utilizzare il sequenziamento di tipo Shotgun sui campioni umani antichi è la facilità con cui è possibile generare, durante l'assemblaggio delle letture, sequenze chimeriche fra DNA contaminante esogeno principalmente batterico, e il DNA endogeno. Per cercare di ovviare a questo problema e per aumentare l'efficienza di recupero delle sequenze endogene di interesse sono stati sviluppati vari metodi di cattura e arricchimento di regioni target del DNA.

I metodi Capture selezionano nella libreria genomica solo i frammenti relativi alla regione d'interesse attraverso l'ibridazione con sonde, amplificandoli e separandoli dal resto della libreria. Le sonde sono costruite sulla base della regione che si intende analizzare e solitamente sono legate ad un adattatore biotinilato che ne permette il recupero dopo l'ibridazione con i frammenti d'interesse. Questo arricchimento in sequenze target presenta alcuni vantaggi: evita di sequenziare anche regioni non collegate all'obiettivo delle analisi (Ávila-Arcos et al. 2011), aumenta notevolmente la profondità di sequenziamento per la regione di DNA target, fornisce una quantità maggiore di informazione sui singoli frammenti per discriminare le sequenze endogene e diminuisce i costi. Esistono diversi approcci di tipo Capture a seconda della regione bersaglio che si vuole studiare. Il metodo di cattura su microchip (*Nimblegen sequence capture*) e quello in fase liquida (*Agilent SureSelect sequence capture*) sono in grado di recuperare fino a 50 Megabasi mentre il metodo PEC (*Primer Extension Capture*) è adatto al recupero di regioni più piccole. Le sonde utilizzate sono costituite da sequenze di DNA complementari alla sequenza bersaglio e possono avere dimensioni variabili, sono costruite per sintesi a partire da desossinucleotidi trifosfati oppure mediante clonaggio in vettori genici. Nel metodo *Nimblegen sequence capture*, la cattura avviene su un supporto solido, un microchip che può essere funzionalizzato con centinaia di migliaia di sonde. Le sonde possono essere disegnate dall'operatore a seconda delle esigenze e anche essere applicate nell'identificazione e isolamento di SNPs, devono quindi essere altamente specifiche. Questo metodo necessita però, di una quantità di DNA substrato dell'ordine di 1-2 microgrammi per effettuare l'ibridazione iniziale, quantità che è quasi impossibile reperire da reperti antichi. L'*Agilent SureSelect sequence capture* cattura in fase liquida, risulta

più adatta a quantitativi di DNA inferiori. Questa metodica utilizza un pool di sonde a RNA biotinilate sintetizzate su un microchip che poi vengono staccate e utilizzate in fase liquida (un buffer di sferette) per ibridare con una library genomica. La fase liquida facilita il legame fra sonde e *libraries* aumentando la resa della cattura (Rizzi et al. 2012), questo permette di utilizzare quantità di DNA substrato inferiori rendendola adatta anche per trattare campioni molto degradati. Nonostante la fase di ibridazione avvenga in condizioni estremamente stringenti e i tutti i lavaggi effettuati, le ibridazioni aspecifiche non possono essere escluse del tutto e quindi non si avrà mai il 100% di ibridazione tra sonde e DNA target. Generalmente si hanno cross-ibridazioni perché la dimensione media dei frammenti di DNA è più lunga di quella delle sonde e i segmenti complementari sulle sonde sono meno della metà di quelli di DNA target. Così restano sempre dei frammenti di DNA non legati in grado di ibridare altri componenti. La PEC, invece, è una metodica molto affidabile se si vogliono recuperare frammenti piccoli: le sequenze d'interesse vengono catturate utilizzando primers biotinilati e estensione con DNA polimerasi. I frammenti ibridizzati sono poi recuperati con biglie ricoperte di streptavidina, sfruttando la sua elevata affinità per la biotina. Questa tecnica è stata utilizzata per studiare i primi genomi mitocondriali neanderthliani e dell'uomo di Denisova.

Il *sistema di cattura Maricic* (Fig.9) deriva dal sistema PEC ma si differenzia da quest'ultimo in quanto le sonde sono costruite home-made a partire dal DNA mitocondriale moderno, con un notevole risparmio in termini economici.

Messo a punto proprio con l'idea di catturare piccoli frammenti, il sistema di cattura Maricic, (Maricic et al. 2010), tsi è anche dimostrato molto efficace e conveniente per lo studio del DNA mitocondriale estratto da reperti antichi. Trova anche impiego in studi forensi e di genetica di popolazione. Questo sistema utilizza sonde costruite a partire da prodotti di PCR che sono in grado di catturare interi DNA mitocondriali. Le sonde sono preparate usando un DNA moderno fresco come stampo per due *PCR long-range*, così da ottenere due prodotti di amplificazione sovrapposti che ricoprono l'intero genoma mitocondriale. I prodotti delle PCR long-range vengono poi sottoposti a sonicazione per ottenere frammenti di una lunghezza compresa fra le 150 e le 300 bp. Successivamente i frammenti sono biotinilati su entrambe le estremità e denaturati, i singoli filamenti ottenuti sono poi immobilizzati su biglie magnetiche rivestite di streptavidina. L'ibridazione con le libraries avviene in fase liquida, incubandole con le biglie per due notti. Poi si effettuano vari lavaggi e infine il DNA arricchito viene eluito e sequenziato con NGS.



**Figura 9:** Protocollo cattura con sistema Maricic. A sinistra: costruzione delle sonde. A destra: costruzione di una libreria indicizzata. In basso: cattura dei frammenti target (mtDNA) sulle biglie, grazie alle sonde.

## 2.2 Criteri di autenticità

Tutte le problematiche e le caratteristiche già elencate, che affliggono lo studio del DNA antico si manifestano poi nella difficoltà di stabilire con certezza assoluta l'autenticità delle sequenze ottenute soprattutto quando si tratta di campioni *sapiens* antichi (Gilbert et al. 2005). I criteri su cui convalidare i risultati si evolvono e si precisano nel tempo insieme alle tecniche sempre più efficienti utilizzate per l'analisi del DNA antico.

Le prime linee guida pubblicate da Pääbo (Pääbo, 1989) riguardavano tre punti:

1. esecuzione di controlli negativi in parallelo per rilevare la contaminazione introdotta da reagenti e soluzione durante l'estrazione
2. eseguire più di un estratto da ogni campione le cui sequenze devono

essere fra loro identiche

3. verificare l'esistenza di una correlazione inversa tra efficienza di amplificazione e dimensione dell'amplicone, che riflette la degradazione del DNA del campione. Questi criteri si sono evoluti in un elenco più ampio e dettagliato di requisiti fino alla pubblicazione dell'articolo *Ancient DNA do it or not at all* (Cooper e Poinar, 2000), dove si elencano tutti gli accorgimenti necessari quando si effettua uno studio sul DNA antico utilizzando la metodologia classica, per ottenere un risultato autentico e affidabile. Le norme stabilite sono riassunte in 9 punti, i ben noti *Golden Criteria*:
  1. Il laboratorio in cui si effettuano le analisi deve essere dotato di aree fisicamente separate una dedicata alle fasi di lavoro pre-amplificazione e una dedicata alle fasi post-amplificazione. (vedi figura)
  2. Allestire controlli negativi sia in fase di estrazione che in fase di amplificazione, così da monitorare l'eventuale presenza di contaminanti dei reagenti e del laboratorio.
  3. Allestire diverse estrazioni e amplificazioni dello stesso campione così da verificare la riproducibilità dei dati ottenuti e monitorare costantemente la contaminazione nelle due fasi. Infatti dato il basso numero di copie in cui si trova il DNA antico non basta una sola estrazione o una sola aliquota dello stesso per dichiarare che il campione è privo di DNA. Generalmente tre estrazioni sono considerate un numero ragionevole prima di abbandonare un campione. Diverse amplificazioni sono importanti per il problema delle misincorporazioni.
  4. Clonare i prodotti di amplificazione, per ogni campione vengono analizzati circa 10-15 cloni per tratto di genoma, con una media di 15 cloni per campione. Il sequenziamento di cloni multipli permette di controllare tutte le differenze presenti fra i prodotti di PCR e permette di stimare la presenza di eventuali errori e il rapporto fra sequenze endogene e contaminanti.
  5. L'analisi del campione deve essere replicata in un altro laboratorio e il risultato deve essere riproducibile. Questa è una precauzione in più per controllare la presenza di eventuali contaminanti presenti sfuggiti ai controlli negativi in fase di estrazione e amplificazione.
  6. Valutare il grado di preservazione del campione facendo analisi biochimiche su altre biomolecole la cui sopravvivenza sia in correlazione con quella del DNA. Queste analisi servono sia da supporto per il risultato, infatti ci aspettiamo di trovare un DNA ben conservato in campioni che mostrano buona conservazione biochimica, sia come procedura di screening iniziale per valutare il grado di conservazione del DNA. Ad esempio per ogni campione è determinato il grado di racemizzazione di tre amminoacidi

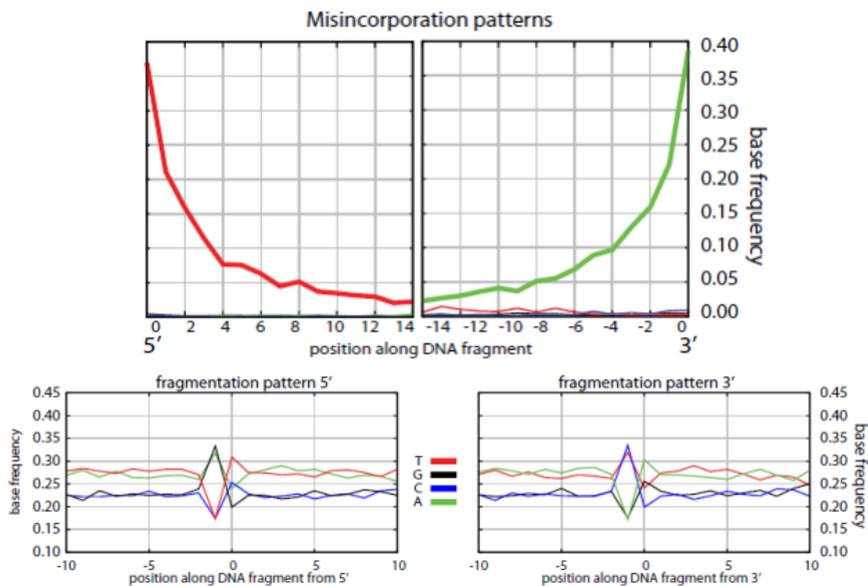
(aspartato, glutammato e alanina), se il valore risulta basso ci si può aspettare una molecola di aDNA ben preservata.

7. Quantificare mediante Real-time PCR o PCR competitiva il numero di molecole amplificabili presenti nell'estratto, in modo da poter prevedere se sono sufficienti o meno per le fasi analitiche successive. Quando la quantità di partenza è bassa (un numero minore di 1000 molecole), non si può comunque escludere la presenza di contaminazione sporadica, soprattutto se si tratta di un campione di DNA umano.
8. Amplificare frammenti di piccole dimensioni fino a un massimo di 200 bp, nell'analisi dei reperti antichi si osserva infatti una correlazione inversa fra efficienza di amplificazione e lunghezza dell'amplificato.
9. Valutare il grado di conservazione del campione analizzando reperti provenienti dallo stesso sito di ritrovamento, così da avere informazioni sulle contaminazioni ambientali.

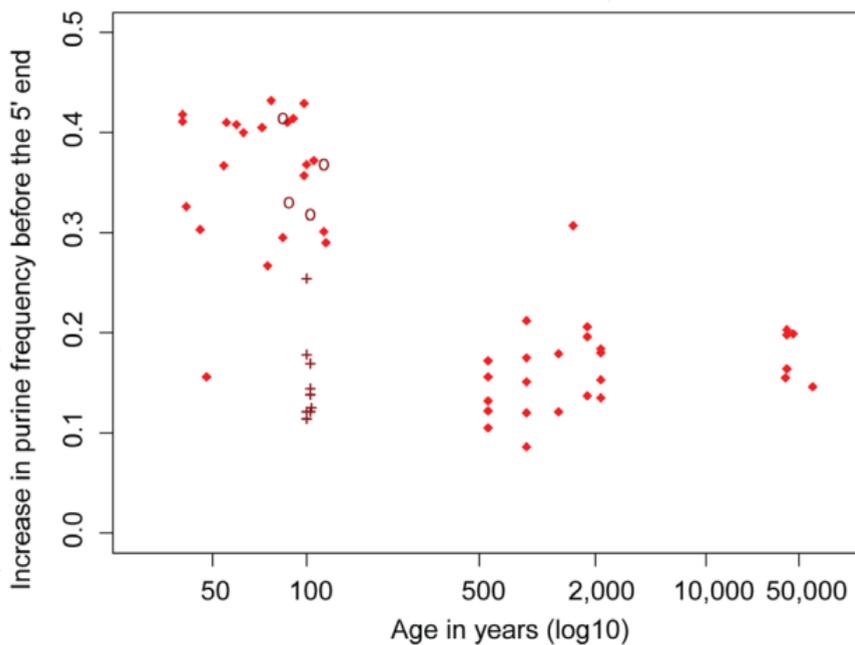
Tuttavia pur seguendo fedelmente tutti criteri e le norme elencate, quando si lavora con la metodologia classica non si può mai avere la certezza completa di aver ottenuto sequenze autentiche. Questo accade soprattutto quando si lavora con campioni antichi di *Homo sapiens*, i quali presentano spesso varianti genetiche che ricadono precisamente nella variabilità delle popolazioni attuali, rendendo impossibile discriminare con sicurezza fra aDNA endogeno e DNA moderno.

Lo sviluppo delle tecnologie di nuova generazione, le caratteristiche dei nuovi metodi di sequenziamento e l'enorme quantità di dati ottenuti per campione ad ogni singola corsa hanno reso necessaria una revisione dei criteri di autenticità già condivisi (Cooper e Poinar 2000; Hofreiter et al. 2001, Pääbo et al. 2004) e di inserirne nuovi, per evitare una errata interpretazione dei dati ottenuti (Green et al. 2009; Krause et al 2010a). I sequenziamenti *high-throughput* offrono nuovi parametri di discriminazione fra sequenze endogene e contaminanti moderne. Prima di tutto, questi nuovi metodi sono in grado di generare sequenze relative a frammenti di DNA di varia lunghezza, basandosi sull'osservazione che il DNA antico è degradato ed è presente in frammenti che vanno dai 50 fino a un massimo di 200 nucleotidi, si può ipotizzare che i frammenti più lunghi siano contaminanti moderni. In secondo luogo, restituiscono letture (*reads*) di entrambe le sequenze dei singoli filamenti di orientamento noto (sequenziamento in *paired-end*) e quindi è possibile osservare la composizione nucleotidica alle estremità 5'-3'. Anche questa caratteristica ci offre uno strumento per distinguere fra sequenze endogene e contaminanti, dato che come già spiegato l'aDNA è caratterizzato da uno specifico pattern di misincorporazione dovuto ai danni *post-mortem* a cui va incontro. La caratteristica diagnostica principale è il pattern di sostituzione C→T che

aumenta verso l'estremità 5' della sequenza e che nella maggior parte dei casi corrisponde ad un pattern complementare G→A all'estremità 3' causata dalla riparazione enzimatica (Fig.10) (Krause et al. 2010; Green et al. 2009). Questo pattern di sostituzione è attribuito, come già spiegato (capitolo 2, *lesioni di miscoding*) alla deaminazione della citosina e tende ad aumentare nel corso del tempo (Fig.11). Quindi il fenomeno risulta più marcato tanto maggiore è l'età del reperto da cui si estrae il DNA, sempre considerando gli altri fattori esterni che influenzano la conservazione della molecola di DNA. Fatte queste considerazioni, la possibilità di analizzare il pattern di composizione al 5'-3' ci permette di capire se una determinata sostituzione nucleotidica è endogena del campione oppure è dovuta ai danni *post-mortem* subiti dal DNA. Infine mediante la sovrapposizione di più *reads* in una regione genomica di riferimento è possibile osservare la prevalenza di purine adiacenti ai siti di frammentazione (Meyer et al. 2012), basandosi sulle analisi che hanno individuato la depurinazione come causa principale della frammentazione (Briggs et al. 2007), questo può essere un altro criterio su cui basare l'autenticità delle sequenze endogene. Secondo il lavoro di Briggs e colleghi (Briggs et al. 2007), lunghezza del frammento, pattern di misincorporazione al 5'-3' e siti di frammentazione in corrispondenza di purine, sono criteri con cui si può discriminare il DNA esogeno da quello endogeno, almeno in alcuni campioni del tardo pleistocene. Mentre uno studio più recente (García-Garcerà, Gigli et al. 2011) ha dimostrato che lunghezza del frammento e composizione in basi in corrispondenza dei siti di frammentazione non possano essere relazionati in modo significativo all'età del campione. Il lavoro ha mostrato infatti che il trattamento del campione con varichina produce in esso un pattern di frammentazione compatibile con quello dei campioni antichi dovuto alla depurinazione. Invece il pattern di misincorporazione è la caratteristica più adatta da mettere in correlazione con l'età del campione (Sawyer et al. 2012). Comunque per valutare questi criteri nei dati generati da sequenziamento NGS, è fondamentale l'utilizzo di strumenti bioinformatici, con script disegnati appositamente per l'analisi dell' aDNA, in grado di definire l'autenticità delle sequenze ottenute.



**Figura 10:** In alto *pattern di misincorporazione* in cui è mostrato che la maggioranza delle sostituzioni C→T si trova all'estremità 5' dei frammenti di aDNA e che poi si traduce in una maggioranza di di sostituzioni G→A all'estremità 3'(krause et al. 2010)



**Figura 11:** La figura mostra l'aumento della frequenza di C→T all'estremità 5' con il passare del tempo.(Sawyer et al.2012)

### 3.SCOPO DEL LAVORO

Nel 2009 un osso frontale umano emerso dai sedimenti fluviali è stato ritrovato lungo una spiaggia del fiume Po in corrispondenza della confluenza con il fiume Adda. Il reperto è stato affidato ai curatori del vicino Museo Paleoantropologico del Po. Nonostante la frammentarietà del reperto è apparso subito chiaro che il frontale presentava caratteristiche di arcaicità non attribuibili all'uomo moderno. Nel corso degli anni, nella stessa zona, sono stati ritrovati numerosi resti di fauna tipicamente glaciale che hanno permesso di ipotizzare la presenza di una giacitura primaria di fossili glaciali localizzata molto probabilmente nella anse del fiume Adda. Tuttavia fino alla scoperta di questo osso frontale non c'erano indicazioni di una frequentazione umana della aree padane da parte di ominidi arcaici quali nello specifico i Neandertaliani. Data l'unicità del reperto e la mancanza di contestualizzazione si è reso pertanto necessario avviare un progetto di studio multidisciplinare basato su metodologie all'avanguardia per arrivare a una descrizione approfondita del reperto e a una sua corretta classificazione.

In questa tesi verranno descritte le analisi genetiche effettuate su una piccola porzione di materiale osseo al fine di caratterizzare il DNA mitocondriale del reperto. Tale marcatore si è rivelato infatti uno strumento molto utile per classificare le diverse forme umane lungo la linea evolutiva degli ominidi, in quanto i range di variabilità riscontrati nelle diverse popolazioni arcaiche non sono sovrapponibili tra di loro e con la variabilità moderna (Briggs et al. 2009; Reich et al. 2009; Meyer et al. 2012). Per effettuare le analisi paleogenetiche è stato utilizzato un approccio di nuova generazione basato su: i. la preparazione di librerie genomiche; ii. la cattura dell'intero genoma mitocondriale tramite ibridazione con sonde biotinilate; iii. il sequenziamento NGS (Next Generation Sequencing); iv. l'analisi delle sequenze ottenute tramite pipeline bioinformatiche specifiche per valutare e massimizzare il livello di autenticità dei dati ottenuti.

Data la complessità del reperto, per il quale non è ricostruibile la storia tafonomica ma che è sicuramente stato sottoposto a forti processi degradativi e a numerose fonti di contaminazione con DNA esogeno, nel corso dello studio si è reso necessario ripetere le analisi utilizzando procedure sia molecolari che bioinformatiche più stringenti rispetto a quelle inizialmente individuate. Per fare questo ci si è avvalsi anche della collaborazione con il Department of Evolutionary Genetics del Max Planck Institute of Evolutionary Anthropology di Leipzig, Germania, in particolare del gruppo di Advanced DNA Sequencing Techniques guidato da Matthias Meyer, presso cui sono state svolte parte delle analisi. Il gruppo di MM è infatti specializzato nella messa a punto di protocolli di estrazione e tipizzazione di DNA da campioni altamente degradati in cui viene massimizzato il recupero delle molecole più corte (<40bp) nonché la successiva analisi bioinformatica

volta a filtrare le sequenze autentiche in base alla presenza di specifici pattern di degradazione (Meyer et al. 2014).

I risultati ottenuti dalle analisi paleogenetiche sono stati infine interpretati anche alla luce dei risultati disponibili per le altre indagini effettuate in particolare di tipo morfologico.

## **4. MATERIALI E METODI**

### **4.1 Campionamento**

Il reperto fossile denominato “Pàus” è attualmente conservato presso il Museo Paleoantropologico del Po di San Daniele Po (CR). È costituito da un osso frontale umano appartenente ad un individuo adulto. Il frontale è quasi del tutto completo ma danneggiato anteriormente in corrispondenza delle regioni glabellare e orbitale.

Parallelamente alle analisi genetiche, la morfologia di Pàus è stata studiata, tramite tecniche di morfometria geometrica, presso il laboratorio di Paleoantropologia e Bio-archeologia della Sapienza Università di Roma in collaborazione con l'International Centre for Theoretical Physics (ICTP) di Trieste. Questa tecnica prevede un approccio quantitativo mediante l'utilizzo delle misure lineari dei tratti morfologici. Si misurano e si confrontano tra loro lunghezze, larghezze e distanze tra particolari punti (Landmark) di determinati reperti per comparare gli individui, siano essi vegetali, animali e come in questo caso, l'osso frontale di un ominide arcaico. La morfometria geometrica è quindi un insieme di metodi che si propongono di analizzare le differenze tra le forme biologiche catturando la geometria complessiva delle strutture oggetto di studio. Per analizzare il reperto di Paus sono state acquisite una configurazione di 6 punti anatomici e una curva di semi-landmark lungo il profilo medio sagittale tra i punti anatomici: bregma e punto più profondo nella regione sopratorale. Le omologhe configurazioni (semi-landmark) sono state acquisite su un campione di confronto piuttosto ampio (N=47) costituito da esemplari appartenenti alle seguenti specie: *Homo ergaster* (N=1), *Homo erectus* (N=6), *Homo heidelbergensis* (N=6), *Homo neanderthalensis* (N=9) e *Homo sapiens* (N=25). Le configurazioni sono state sottoposte ad Analisi generalizzata di Procruste e a Sliding (Gunz, 2005). Dai risultati delle analisi di morfometria geometrica si evince come la morfologia di Pàus sia compatibile con il range di variabilità riscontrato all'interno del sottocampione neanderthaliano e mai sovrapposta a quella delle altre specie di confronto. (Fabio Di Vincenzo, Giorgio Manzi, Antonio Profico comunicazione personale).

## **4.2 Ambiente di lavoro e prevenzione dalle contaminazioni**

Il problema principale nello studio del DNA antico è la contaminazione da DNA moderno, per questo motivo, entrambi i laboratori coinvolti nella analisi molecolari su Paus sono organizzati secondo specifici criteri e operano seguendo una serie di norme atte a limitare al massimo questo rischio. In generale ogni laboratorio è strutturato in due settori fisicamente separati e dedicati alle diverse fasi del lavoro sperimentale: la zona di *Pre-Amplificazione (Pre o Clean Room)* e la zona di *Post-Amplificazione (Post)*. Nel *Pre* si effettua tutta la fase sperimentale precedente all'amplificazione del DNA: pulizia e polverizzazione dei reperti ossei, estrazione del DNA, allestimento della miscela di reazione per la PCR e alcune fasi della preparazione della libreria genomica. Nel *Post* invece, avvengono tutte le fasi di lavoro che prevedono DNA già amplificato e non più a rischio di contaminazione da DNA moderno quali controllo elettroforetico della PCR, clonaggio, reazioni di sequenza, fasi finali del flusso di lavoro Illumina e cattura del DNA mitocondriale. Per mantenere la sterilità del *Pre*, vi si può accedere solo se provvisti di guanti, tuta, camice cuffia mascherina e copri scarpe. L'area del *Pre* è dotata di un sistema di ricircolo dell'aria e viene regolarmente decontaminata mediante pulizia delle superfici di lavoro con ipoclorito di sodio e irraggiamento con radiazioni UV a 254 nm. Tutte le superfici di lavoro, utilizzate per le fasi di lavoro dove è massimo il rischio di contaminazione (pulizia, estrazione, preparazione della mix per la PCR, prime fasi librerie), vengono usualmente lavate con ipoclorito di sodio e irraggiate con radiazioni UV, prima e dopo l' utilizzo. Anche gli strumenti, altri materiali e la vetreria, dove possibile sono costantemente sterilizzati con ipoclorito di sodio o acido acetico al 30% (per materiali in metallo) e raggi UV. I reagenti o uno qualsiasi degli strumenti del *Post* non vengono mai introdotti nel *Pre* e così gli operatori, che nell'arco della stessa giornata, anche se sono semplicemente entrati nel *Post*, non possono più entrare nella zona *Pre*. Particolare attenzione è adottata per tutti i reagenti utilizzati nel *Pre* che vengono aliquotati, cercando di mantenere le loro condizioni sterili, evitando così l'uso ripetuto da una singola stock. Durante le varie fasi di lavoro vengono analizzati i controlli negativi, contenenti solo la miscela dei reagenti, sia per l'estrazione che per la libreria genomica, così da poter monitorare costantemente la presenza/assenza di contaminanti.

## **4.3 Pulizia e polverizzazione dei campioni**

Sia la pulizia del reperto che la sua polverizzazione sono state effettuate nel laboratorio *Pre*, sotto cappa aspirante, lavata con ipoclorito di sodio e sterilizzata sotto i raggi UV prima e dopo l'utilizzo. In generale, una volta che il campione giunge in laboratorio deve essere accuratamente pulito prima di procedere con la polverizzazione, al fine di eliminare dalla sua

superficie eventuali residui di terreno, polvere o materiale biologico, come cellule epiteliali che possono essersi deposte durante lo scavo o la successiva manipolazione da parte degli operatori. Considerando che il DNA contaminante può permeare la superficie delle ossa fino a uno spessore di 1-2 mm questo spessore superficiale è stato rimosso usando un trapano odontoiatrico con microfresse in titanio. Successivamente, il frammento osseo pulito viene collocato nel crosslink (Microlink, PBI, Italia), all'interno del quale ogni superficie del frammento viene irradiata ortogonalmente con raggi UV a 254 nm per 45 minuti in questo modo l'eventuale Dna eterologo è reso inattivo. A questo punto si può procedere con la polverizzazione: usando lo stesso trapano odontoiatrico con microfresse viene prodotta la polvere necessaria per l'estrazione praticando un prelievo minimamente invasivo. La zona dell'osso in cui si effettua il prelievo è accuratamente scelta, al fine di mantenere inalterate le dimensioni dell'osso e non interferire con eventuali rilievi antropometrici e studi morfometrici successivi. Per Paus sono stati ottenuti circa 100 µg di polvere d'osso.

#### **4.4 Estrazione del Dna da polvere d'osso**

Come processo generale l'estrazione del DNA consiste in tre fasi: la prima vede la rottura delle membrane cellulari mediante solubilizzazione progressiva delle strutture cellulari che lo contengono, denaturazione delle proteine e inattivazione di enzimi che potrebbero degradarlo, mediante il trattamento con soluzioni contenenti tensioattivi, proteasi e agenti chelanti. La seconda fase è quella della *deproteinizzazione* nella quale si ha la rimozione delle proteine cellulari ed extracellulari grazie a proteasi che le idrolizzano o le denaturano. Nell'ultima fase il DNA puro è separato dalle altre componenti cellulari. L'estrazione si effettua nel *Pre*, sotto la stessa cappa usata per pulizia e polverizzazione, indossando guanti chirurgici e nel rispetto di tutte le norme già elencate per assicurare l'assoluta sterilità del procedimento. L'estrazione è un passaggio critico, contaminazioni in questa fase di lavoro inficiano il proseguimento e il successo di tutto lo studio, per questo durante tutti i passaggi del lavoro è necessario includere un controllo negativo (K- estrazione), la cui successiva analisi permette di visualizzare se gli estratti sono contaminati o meno. I protocolli per estrazione più utilizzati nello studio del DNA antico si basano sull'utilizzo di particelle di silice, un materiale che presenta elevata affinità per gli acidi nucleici a pH bassi e in presenza di sali caotropici.

Un protocollo prevede l'utilizzo della Guanidina Tiocianato e della silice in soluzione per trattenere le molecole di acido nucleico (Rohland e Hofreiter 2007), il secondo utilizza la guanidina idrocloride e colonnine che immobilizzano la silice su filtri (kit MinElute Qiagen) (Dabney et al. 2013). Per questo lavoro sono stati usati entrambi i protocolli su due diverse aliquote di polvere di osso di 150 e 25 µg rispettivamente.

Il primo protocollo prevede quattro fasi di lavoro articolate in due giorni consecutivi. La prima fase ha lo scopo di digerire la polvere d'osso e rilasciare il DNA attraverso l'incubazione con un volume di circa 5ml di **EXTRACTION BUFFER** che contengono: 360 µl di **EDTA 0.5 M a pH 8**, che chela gli ioni  $\text{Ca}^{2+}$  favorendo la decalcificazione della componente inorganica della matrice ossea; 47.2 µl di **Proteinasi K** (PK) 18 ng/µl, enzima proteolitico in grado di degradare la componente proteica; 170 µl di SSC 20X a pH 7.2; 170 µL SDS 10%, detergente tensioattivo anionico, permette la rottura delle membrane, il tutto portato a volume con  $\text{H}_2\text{O}$  di grado HPLC. Dopo l'incubazione *over night* i campioni vengono centrifugati per due minuti a 3300 g, così da separare il surnatante con il DNA in soluzione dal pellet in cui si deposita la componente che non è stata digerita. Il pellet di polvere restante viene conservato in congelatore per eventuali nuove analisi. Il surnatante viene trasferito in una nuova falcon sterile e incubato con 4 ml di **BINDING BUFFER** e 100 µl di sospensione silica. Il Binding Buffer è composto (per singolo campione) da: 2,056 g di Guanidina Tiocianato, 0,35 ml di Sodio Acetato 3M, 1,61 ml di  $\text{H}_2\text{O}$  HPLC; si prepara fresco il giorno prima dell'uso, ma si può anche preparare da una settimana all'altra senza però superare il limite temporale massimo di un mese. Questa soluzione quasi satura di guanidina a PH acido favorisce l'adsorbimento del DNA alle particelle di silica. Le falcon dopo essere state sigillate con Parafilm vengono lasciate nel rotore al buio a temperatura ambiente per tre ore, al termine delle quali vengono di nuovo centrifugate a 3300 g per 2 minuti e il surnatante scartato; il pellet di silica con il DNA viene risospeso in 1 ml di Binding Buffer e trasferito in nuove eppendorf da 2 ml che vengono poi centrifugate alla massima velocità (1400 rpm) per 10 secondi. Una volta rimosso il surnatante, il pellet recuperato è sottoposto a lavaggi così da eliminare qualsiasi molecola diversa dal DNA. I lavaggi si effettuano usando **Wash Buffer**, per 50 ml di soluzione si usano: 26,65 ml di etanolo al 100%, 1,25 ml di NaCl 5M, 0,5 ml di Tris a PH 8 1 M e  $\text{H}_2\text{O}$  per portare a volume. Il pellet è sospeso in 1ml di Wash Buffer e centrifugato per 10 secondi alla massima velocità (14000 rpm), questo step si ripete due volte. Al termine del secondo lavaggio si elimina completamente il surnatante e il pellet di silica con l'estratto adsorbito viene fatto asciugare per 15 minuti a temperatura ambiente così da consentire la completa evaporazione dell'etanolo. L'ultima fase è quella dell'Eluizione, il pellet viene incubato con 50 µl di TET composto da EDTA 0,5 M, Tris 1M entrambi a pH 8, Tween 20 e  $\text{H}_2\text{O}$  per portare a volume, creando così un ambiente adatto al distacco delle molecole di DNA dalla Silica e il loro rilascio in soluzione, dopo un passaggio in centrifuga per 1 minuto a velocità massima, il surnatante: DNA estratto, viene trasferito in nuove eppendorf e può essere usato direttamente oppure conservato in freezer a  $-20^\circ\text{C}$ . Il secondo protocollo si articola in due giorni di lavoro e si avvale della combinazione di due kit di purificazione, lo Zymo-Spin V Column, Zymo Reserch e il Min Elute. Tutti i

reagenti sono preparati in falcon da 50 ml e sterilizzati nel cross-link a 7000J per 45 minuti con il tappo del falcon aperto prima di utilizzarli. La PK è aggiunta all'**EXTRACTION BUFFER** subito dopo l'estrazione. Prima di procedere con l'estrazione i Reservoir del kit Zymo vengono lavati con acqua e varichina e poi sterilizzati nel cross-link, insieme a tutti i falcon e le eppendorf necessarie per il procedimento. 25 mg di polvere d'osso vengono posti in eppendorf da 2ml e incubati con 1ml di EXTRACTION BUFFER per 10ml di soluzione si usano: 745 µl di H<sub>2</sub>O, 9 ml di 0,5 M a pH8, 250 µl di PK 10 mg/ml, 5 µl di Tween 20. I tubini vengono vortexati fino a che non si nota più nessun pellet sul fondo o sulle pareti, poi sigillati con parafilm e lasciati sul rotore a 37°C *overnight*. Il giorno successivo per ogni campione viene preparato un falcon da 50 ml con 10 ml di **BINDING BUFFER** e 400 µl di Sodio Acetato 3M. Per 10 ml di BINDING BUFFER si usano: 23.88 g di Guanidina idrocloride 5M, H<sub>2</sub>O fino a 30 ml, isopropanolo fino a 50 ml e 25 µl di Tween 20. I falcon contenenti i campioni e l'**EXTRACTION BUFFER** vengono centrifugati per 2 minuti alla massima velocità così da compattare tutti i residui di polvere non digeriti. Il supernatante viene trasferito nel falcon contenente il BINDING BUFFER e Sodio Acetato mentre il pellet viene conservato per estrazioni successive.

A questo viene costruito l'apparato Reservoir(Zymo research)/MinElute: il Reservoir viene forzato nell'apertura della colonnina MinElute e poi collocato in un falcon sterile da 50 ml (Fig.12).



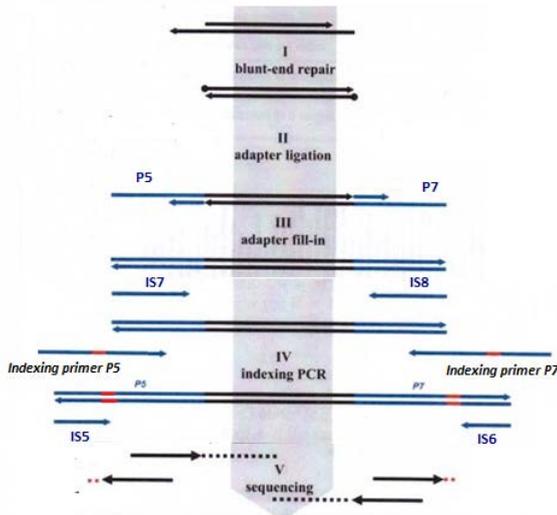
**Figura 12:** composizione dell'apparato Reservoir-Zymo/MinElute

Prima di procedere gli apparati vengono sottoposti a un test di stabilità: 4 minuti in centrifuga a 1500 rpm, se il test va a buon fine si procede con lo step successivo. Il campione con il BINDING BUFFER viene versato nell'apparato Reservoir/MinElute e messo in centrifuga 4 minuti a 1500 rpm. A questo punto il DNA è rimasto nella resina della colonnina MinElute, l'apparato viene rimosso dal falcon e collocato nel collection tube del MinElute, poi con molta attenzione si rimuove il reservoir. Le colonnine MinElute vengono centrifugate 1 minuto a 6000rpm e il liquido percolato eliminato. A questo punto si procede con una normale purificazione MinElute (descritta nel procedimento di preparazione della library) saltando l'aggiunta di PB che è già presente nel BINDING BUFFER. Alla fine si eluisce in 25  $\mu$ l di TET aspettando 5 minuti prima di centrifugare. L'eluizione in TET viene ripetuta per due volte quindi alla fine abbiamo 50  $\mu$ l di estratto da trasferire in una nuova eppendorf (siliconata).

#### **4.5 Preparazione della libreria a doppio filamento (DS-Library)**

Le librerie si preparano secondo un protocollo multiplex modificato, pubblicato da Meyer e Kircher (Meyer e Kircher 2010) (Fig.13), di ogni estratto compreso il controllo negativo dell'estrazione e un controllo negativo della libreria portato avanti fino all'ultima fase, cattura compresa. Tutte le fasi precedenti dalla pulizia dei campioni all'estrazione, fino all'amplificazione della library stessa si svolgono nel *Pre*, sottostando a tutte le norme descritte per ridurre al minimo le contaminazioni e mantenere l'assoluta sterilità.

Il protocollo si articola nelle seguenti fasi: Riparazione dei danni all'estremità dei frammenti estratti (Blunt-end repair), ligazione degli adattatori (Adapter Ligation), riparazione degli eventuali nick presenti fra adattatori e inserto di DNA (Fill-in), indicizzazione (Indexing PCR) e arricchimento (Accuprime PCR).



**Figura 13:** Schema riassuntivo degli step previsti per la preparazione della libreria a doppio filamento (Meyer e Kircher 2010)

#### 4.5.1 Blunt-end repair

Questa prima fase di preparazione della libreria prevede la riparazione di eventuali sporgenze a singolo filamento alle estremità 3'-5' dei frammenti estratti, per prepararli al successivo legame con gli adattatori. La reazione di riparazione può essere rimozione o ricostruzione della sequenza avviene ad opera della T4 polimerasi e della T4 polinucleotide chinasi (PNK): la PNK catalizza la formazione di un gruppo fosfato all'estremità 5' del frammento mentre la T4 polimerasi guida l'estensione delle estremità a singolo filamento fino a generare estremità piatte (*Blunt-end*) adatte al legame con gli adattatori. La mix di reazione è preparata in tubini per PCR da 0,2 ml, per un campione (volume finale 50 µl) e così composta: 5 µl NEB Buffer 2 (stock conc. 10X final conc. 1X); 2 µl dNTP mix 2,5 mM (stock conc. 2500 µM, final conc. 100 µM); 4 µl BSA (stock conc. 10 mg/ml, final conc. 0.8 mg/ml); 5 µl ATP 10 mM, final conc. 1 mM); 2 µl T4 PNK (stock conc. 10 U/µl, final conc. 0,4 U/ µl); 0,4 µl T4 Polymerase (stock conc. 3 U/ µl, final conc. 0,024 U/ µl); 20 µl di DNA template; H<sub>2</sub>O fino a volume.

I tubini vengono caricati in un termociclatore impostato con il seguente profilo termico:

- 15°C per 15 minuti
- 25°C per 15 minuti

Il prodotto ottenuto viene purificato utilizzando il MinElute Purification Kit (Quiagen).

Ad ogni campione si aggiungono 200 µl di PB buffer e si caricano nell'apposita colonnina MinElute. Si centrifuga a 14000 rpm per 1 minuto e si scarta il liquido ottenuto. Poi si aggiungono 600 µl di PE buffer alla

colonnina e di nuovo centrifugazione a 14000 rpm per 1 minuto. L'eluato viene nuovamente scartato e le colonnine ruotate di 180° prima di centrifugarle nuovamente 1 minuto a 14000 rpm, in questo modo si eliminano i residui di PE in tutti i punti del filtro della colonnina. Le colonnine vengono trasferite in nuove eppendorf siliconate da 1,5 ml sterili e si aggiungono 18 µl di TET nel centro del filtro. Si attende 5 minuti e poi si passano in centrifuga 1 minuto a 14000 rpm. L'eluato, DNA Blunt-end repair, è trasferito in nuove eppendorf da 1,5 ml siliconate e sterili ed è pronto per la fase successiva.

#### 4.5.2 Adapter ligation

In questa fase si legano all'estremità della molecola di DNA gli adattatori Solexa Illumina specifici per il sequenziamento. Gli adattatori Solexa Illumina sono preparati a partire da tre oligonucleotidi IS1\_adapter\_P5.R, IS2\_adapter P7.F e IS3\_adapter\_P5+P7 costituiti rispettivamente dalle seguenti sequenze: 5'-AATGATACGGCGACCACCGA, 5'-CAAGCAGAAGACGGCATACGA E 3'-AGATCGGAAGAGC. La preparazione della Solexa Adapter Mix prevede la costituzione di due adattatori solo parzialmente a doppio filamento avendo l'oligonucleotide IS3\_adapter\_P5+P7 complementare a solo una porzione degli altri due oligonucleotidi P5 e P7 che danno il nome agli adattatori. Questa conformazione particolare, con una sola estremità piatta, garantisce che gli adattatori si leghino alle molecole di DNA con l'orientamento corretto. Gli adattatori P5 e P7 si preparano separatamente e poi vengono riuniti in un'unica mix: Solexa Adapter Mix. Per la reazione di *Adapter ligation* si utilizza il Quick Ligation Kit NEB, Ipswich, USA. Le quantità per singolo campione in un volume finale di 40 µl sono le seguenti: 20 µl di Quick Ligase Buffer (2X) (stock conc. 2X, final conc. 1X); 1 µl di Solexa Adapter Mix (stock conc. 10000 nM, final conc. 250 nM); 18 µl di DNA Blunt-end o di controllo negativo. 21 µl di Master Mix sono aggiunti ad ogni campione poi ad ogni tubino si aggiunge 1 µl di Quick ligase (stock conc. 5U/ µl, final conc. 250 nM) l'enzima che catalizza il legame degli adattatori, spipettando per mescolare uniformemente e lasciate incubare a temperatura ambiente per 20 minuti. Avvenuta la reazione si procede con purificazione con MinElute Purification Kit eluendo in 20 µl di TET e si passa alla fase successiva.

#### 4.5.3 Adapter Fill-in

Con questo passaggio si estendono le porzioni degli adattatori ancora a singolo filamento, la reazione è operata dalla Bst polimerasi. Per un campione e un volume finale di 40 µl si usano: 4 µl di Isothermal buffer (10X) ( stock conc. 10X, final conc 1X), 2 µl dNTP 2,5 mM (stock conc. 2500 nM, final conc 125 nM); 2 µl Bst polymerase 2.0 ( stock conc. 8 U/ µl, final conc. 0,4 U/µl); 12 µl H<sub>2</sub>O; 20 µl Dna templato. 20 µl Master Mix vengono aggiunti ai 20 µl di DNA ottenuti per ogni campione dall' *Adapter*

*ligation*, la reazione avviene nel termociclatore impostato con il seguente profilo termico:

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

A questo punto vengono prelevati 2 µl di ogni libreria da quantificare mediante Real Time PCR per conoscere la quantità iniziale di DNA e ottimizzare le fasi successive. I 38 µl rimanenti vengono trasferiti in nuove eppendorf siliconate sterili e congelati.

#### **4.5.4 Real-Time PCR (RT-PCR )**

La RT-PCR è una PCR quantitativa (qPCR) che permette grazie, all' uso di un fluoroforo, di monitorare la fase esponenziale della reazione di PCR e quantificare il contenuto di DNA presente in ogni libreria. La mix di reazione è preparata usando il kit sso fast Eva green che utilizza il Syber green come fluoroforo e i primers Solexa Primer Fwd IS8 e Solexa Primer Rvs IS7 specifici per le estremità degli adattatori. Le quantità di reagenti per un singolo campione e un volume finale di 20 µl sono: 5 µl di supermix EvaGreen, 0.5 µl di ogni primer e 3 µl di H<sub>2</sub>O. Insieme ai campioni vengono caricati in doppio degli standard a concentrazione nota e la libreria è generalmente diluita 1:10 e 1:100 per farla ricadere nella retta standard, mentre i controlli negativi sono utilizzati tali e quali. L'apposita plate caricata con librerie e standard viene posta nella macchina (MiniOpticon System Biorad) e si utilizza il seguente profilo termico:

- Attivazione iniziale: 95°C per 15 minuti;

Per 40 cicli:

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

L'impostazione della curva di Melting resta quella del programma standard.

La qPCR ci fornisce la concentrazione di DNA per ogni campione, prima di proseguire con l'*Indexing* è necessario fare alcuni calcoli necessari all' allestimento della reazione. Si crea un foglio di calcolo in Excel in cui vengono riportati i valori ottenuti per tutte le librerie diluite e mediante formule preimpostate si riportano a una concentrazione, si fa una media dei valori così ottenuti e la si rapporta al volume di libreria rimasto (valore iniziale post *Fill-in* meno quantità usata per la RT-PCR), ottenendo così il numero di molecole di DNA totali. Conoscere il numero di molecole di DNA totali è necessario dato che in fase di *Indexing* il valore totale delle copie di DNA deve raggiungere il valore di 10<sup>13</sup>(circa il plateau della PCR) però oltrepassarlo, la quantità di DNA iniziale va regolata dividendo i campioni in più aliquote (split in 2,4,6,8) e regolando il numero di cicli di amplificazione della reazione calcolati utilizzando una tabella con i fattori di amplificazione di ogni ciclo.

#### 4.5.5 Indexing PCR

In questa fase le librerie vengono indicizzate scegliendo diverse combinazioni degli indici disponibili per ogni campione, questo ci permette di sequenziare in unico *pool*, librerie di campioni diversi e successivamente di discriminare un campione dall'altro durante l'analisi delle sequenze. 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 step di amplificazione e quantificazione. Le sequenze degli indici relativi agli indexing-primer P5\_iPCR-LP complementari all' adattatore P5 e agli indexing-primer Sol\_iPCR-MPI complementari all'adattatore P7 sono riportate in figura 14:

P7 primer ID	P7 sequence	P5 primer ID	P5 sequence
Sol_iPCR-MPI-66	ACCTTGC	P5_iPCR-LP-22	GCATTGG
Sol_iPCR-MPI-88	ATAGAAG	P5_iPCR-LP-19	GAATCTC
Sol_iPCR-MPI-83	AGCGCTG	P5_iPCR-LP-14	TTGGATC
Sol_iPCR-MPI-56	AACTGAG	P5_iPCR-LP-10	CCGGTAC
Sol_iPCR-MPI-84	AGTATAC	P5_iPCR-LP-1	TCGCAGG
Sol_iPCR-MPI-92	ATTATCT	P5_iPCR-LP-12	TTGAAGT
Sol_iPCR-MPI-93	ATTCGAC	P5_iPCR-LP-28	AGATAGG
Sol_iPCR-MPI-67	ACGACCT	P5_iPCR-LP-39	CGAATGC
Sol_iPCR-MPI-81	AGCAAGC	P5_iPCR-LP-44	ACTGGAC
Sol_iPCR-MPI-68	ACGATTC	P5_iPCR-LP-17	AGGTACC
Sol_iPCR-MPI-57	AAGAATT	P5_iPCR-LP-42	CGCGCAG
Sol_iPCR-MPI-53	AACGCCT	P5_iPCR-LP-40	TTCGCAA
Sol_iPCR-MPI-59	AAGCTCT	P5_iPCR-LP-45	AGCAGGT
Sol_iPCR-MPI-96	CAACTAA	P5_iPCR-LP-8	GCTCGAA

**Figura 14:** sono mostrate le sequenze degli indici, contenute all' interno degli *indexing primer*, P7 e P5.

La scelta della combinazione degli indici, in grado di caratterizzare ogni campione, è molto importante per evitare la generazione di errori durante l'acquisizione delle immagini nella piattaforma illumina durante il sequenziamento e per l'esatto riconoscimento dei campioni sequenziali nello stesso *pool*. La reazione di indicizzazione è una PCR, i reagenti necessari per singolo campione sono: 10 µl di Buffer 10X (stock conc.10X, final conc, 1X); 1 µl di dNTP mix 25 mM (stock conc. 25 mM, final conc. 0,25 mM); 1,5 µl (stock conc. 20X, final conc 0,3X), 1 µl di Pfu Turbo Polymerase (stock conc. 2,5 U/ µl, final conc, 0,025 U/ µl). Una volta preparata la Master Mix si mette nei tubini da 0,2 ml sterili, ai quali viene poi aggiunto il volume di libray adeguato a raggiungere la soglia di 10<sup>13</sup> molecole di DNA

(calcolato per ogni campione come descritto nel paragrafo precedente) e H<sub>2</sub>O fino a volume 100 µl. Per aumentare la resa della reazione il volume di library ottenuto dalla *Fill-in* viene “splittato” in più aliquote ognuna delle quali viene amplificata cercando di raggiungere la soglia di 10<sup>13</sup> molecole. Ad ogni campione vengono separatamente aggiunti 2 µl di ciascun primer P7 e P5 (stock conc. 10 mM, final conc. 0,2 µl).

Il profilo termico utilizzato per l'Indexing PCR è il seguente:

-Step di attivazione: 95°C per 2 minuti

Per 10 cicli:

-Step di denaturazione: 95°C per 30 secondi

-Step di annealing: 58 °C per 30 secondi

-Step di amplificazione: 72°C per 1 minuto

Dopo 10 cicli

-Step finale di estensione: 72°C per 10 minuti

Conclusa la reazione, le library vengono purificate con MinElute Purification Kit (Qiagen), come descritto in precedenza, durante la purificazione le aliquote delle library fatte per l' indexing vengono riunite utilizzando una sola colonnina ogni 4 aliquote, e si eluisce in 50 µl di TET. 2 µl di ogni library vengono prelevati per eseguire la quantificazione del prodotto di Indexing, mediante Real-Time PCR.

#### **4.5.6 Real-Time PCR post Indexing**

Il kit e il protocollo utilizzati sono gli stessi della RT-PCR eseguita dopo la Fill-In, così come il profilo termico ma le librerie devono essere diluite 1:100 e 1:100 affinché cadano nella retta standard e si utilizzano primers Bridge IS5 e IS6 complementari alla sequenza degli indexing-primers esterna a quella degli indici, gli stessi primers sono utilizzati per la successiva fase di amplificazione, *AccuprimePCR*. La quantità di DNA ottenuto a questo step viene confrontato con quella ottenuta con la RT-PCR Post Fill-in, il rapporto tra i due ci indica un valore di efficienza che deve essere > di 1 (l' ideale è > di 100).

Anche in questa fase sulla quantità di DNA ottenuta è calibrato il numero di cicli e il numero di aliquote da fare di ogni library così che nella reazione di amplificazione successiva non si superari la quantità di 10<sup>13</sup> molecole di DNA, secondo la stessa logica usata per la reazione di *Indexing*.

#### **4.5.7 Accuprime PCR**

In questa fase la quantità di molecole indicizzate viene arricchita, mediante la reazione di amplificazione *Accuprime PCR* usando i primers Bridge IS5 e IS6 . I reagenti necessari per un campione sono : 10 µl di AP Pfx Rnx mix 10X (stock conc. 10X, final conc. 1X); 3 µl di primer IS5 10 µM (stock conc. 10 µM, final conc. 0,3 µM); 3 µl di primer IS6 10 µM (stock conc. 10 µM, final conc. 0,3 µM); 0,8 µl di Accuprime Pfx DNA Polymerase (stock

conc. 2,5 U/  $\mu$ l, final conc. 0,02 U/  $\mu$ l). Poi si aggiunge il volume di library calcolato dalla q-PCR e H<sub>2</sub>O fino a volume 100  $\mu$ l.

Il profilo termico per la reazione è il seguente:

-Attivazione: 95°C per 2 minuti;

per un numero di cicli calcolato sulla base dei risultati della q-PCR precedente:

-Denaturazione: 95°C per 15 secondi;

-Annealing: 60°C per 30 secondi;

-Amplificazione: 68°C per 1 minuto;

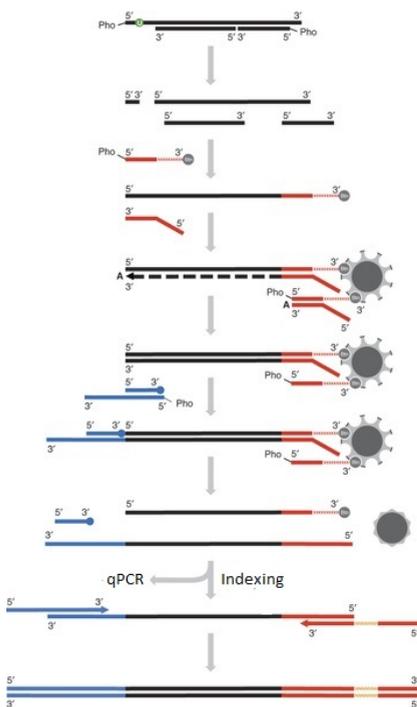
Alla fine dei cicli:

-Amplificazione finale: 68°C per 5 minuti

I prodotti della reazione vengono purificati con il MinElute Purification Kit (Qiagen), e analizzati all'Agilent Bioanalyzer 2100, usando il kit DNA1000, per stimare il profilo qualitativo e quantitativo dei campioni prima di procedere con la cattura del DNA mitocondriale.

#### **4.6 Preparazione della libreria a singolo filamento (SS-Library)**

La libreria a singolo filamento ha come obiettivo principale quello di recuperare ed indicizzare sequenze di DNA anche molto corte, inferiori a 40 pb, che in alcuni casi possono costituire la maggior parte delle molecole endogene di un campione. Il protocollo si articola in due giorni di lavoro: il 1° giorno è articolato nelle seguenti fasi: decontaminazione degli adattatori, Denaturazione e defosforilazione del primo adattatore. Il 2° giorno vede l'allestimento delle seguenti reazioni: Legazione dei prodotti alle biglie magnetiche, Annealing ed estensione dei primers, Ligazione del secondo adattatore ed eluzione finale (Fig.15). Come per la libreria a doppio filamento tutte le fasi di lavoro fino all'indexing compreso vengono svolte nel laboratorio *Pre-PCR*



**Figura 15:** Schema della libreria a singolo filamento

#### 4.6.1 Decontaminazione degli adattatori

La prima fase della libreria a singolo filamento prevede la decontaminazione dell'adattatore CL78-7 e viene preparata periodicamente.

Per una reazione, in un volume finale di 22,8 ul, aggiungere 17ul di adattore CL78-7 (10uM), 2,3 ul di CircLigase Buffer II (10x), 1,2 ul di  $MnCl_2$  (50mM). Vortexare e centrifugare brevemente, incubare nel thermal cycle a 95°C per un minuto e poi mettere immediatamente in ghiaccio. Aggiungere 2.3  $\mu$ L Exonuclease I, centrifugare brevemente, incubare per 20 minuti a 37°C e poi 1 minuto a 95°C per disattivare l'enzima.

#### 4.6.2 Denaturazione e defosforilazione del primo adattatore

In un volume finale di 42ul per ogni campione aggiungere 12,5 ul di acqua HPLC, 8 ul di CircLigase Buffer II (10x), 4 ul di  $MnCl_2$  (50mM), 2 ul di Tween-20 (2%), 0,5 ul di User enzime se richiesto e 15 ul di DNA. Oltre ai campioni mettere anche il controllo positivo rappresentato da CL104 per un miglior controllo del risultato.. Incubare per 1 ora a 37°C. Se l'impiego dell'User enzyme non è richiesto, nella mix di reazione sostituirlo con acqua HPLC e saltare l'incubazione. Aggiungere 1 ul di FastAP (1 U /  $\mu$ l) per ogni reazione ed incubare 10 minuti a 37°C e poi 2 minuti a 95°C, mettere immediatamente in ghiaccio.

Per raggiungere un volume finale di 80 ul aggiungere alla reazione

precedente 31,7 ul di PEG-4000 (50%), 1,3 ul di 10  $\mu$ M CL78-7 (Cf. 4.6.1), 4 ul di CircLigase II (100 U /  $\mu$ l). Incubare per 1 ora a 60°C e portare poi a 4 °C , aggiungere EDTA (0,5 M) e conservare a -20°C . Di solito a questo punto termina il primo giorno di lavoro in laboratorio.

#### **4.6.3 Ligazione dei prodotti alle biglie magnetiche**

Sospendere le biglie magnetiche (MyOne C1) e per ogni campione trasferire 20 ul delle stesse in un tubino da 1,5 ml. Lavare le biglie due volte con 500 ul di buffer appositamente preparati (1xBWT+SDS). Sospendere le biglie in 250 ul di 1xBWT+SDS e trasferirle in un nuovo tubino da 1,5 ml. Incubare la reazione fatta il giorno precedente (Cf. 4.6.2) per 1 minuto a 95°C ed immediatamente mettere in ghiaccio. Aggiungere alla reazione le biglie. Incubare a temperatura ambiente sul rotore per 20 minuti. Utilizzando il rack magnetico separare le biglie e recuperare la parte liquida traferendola in un tubino nuovo, aggiungendo poi alle biglie 200 ul di buffer 0.1xBWT+SDS e trasferire nel ThermoShaker impostato su 25 °C e con agitatore ogni 8 secondi. Aggiungere 200 ul di buffer  $\mu$ l 0.1x BWT e trasferire nuovamente nel ThermoShaker con le impostazioni precedenti.

#### **4.6.4 Annealing ed estensione dei primers**

Ai tubini messi da parte come descritto nel paragrafo 4.6.3, aggiungere 39,1 ul di acqua HPLC, 5 ul di buffer Klenow (10x), 0,4 ul di dNTPs (25mM), 2,5 ul di Tween-20 (1%), 1ul di CL130 (100  $\mu$ M). Centrifugare brevemente, scartare il supernatante e sospendere le biglie nella reazione descritta nel paragrafo 4.6.2. Trasferire nel termal cycle ed incubare per 2 minuti a 65°C ed immediatamente mettere in ghiaccio. Aggiungere 2 ul di frammenti di Klenow (10U/ $\mu$ l). Trasferire i campioni nel ThermoShaker settato a 25°C ed agitatore ogni 3 secondi secondo il seguente programma impostato: Step 1 5 minuti, 25°C, 1500 Hz; Step 2 25 minuti, 35°C, 1500 Hz.

Utilizzando il rack magnetico scartare il surnatante. Sospendere le biglie in 200 ul di buffer 0.1xBWT+SDS, trasferire i campioni nel ThermoShaker settato a 25°C ed agitare 8 secondi. Centrifugare brevemente ed eliminare il surnatante. Sospendere le biglie in 100 ul di buffer Stringency-Wash, trasferire nel ThermoShaker secondo il seguente programma: 3 minuti, 45°C, 1500 Hz. Utilizzando il rack magnetico scartare il surnatante e sospendere le biglie in 200 ul di buffer 0.1x BWT, trasferire i tubi nel rack magnetico.

#### **4.6.5 Ligazione del secondo adattatore ed eluizione finale**

Centrifugare brevemente la reazione contenete le biglie (Cf. 4.6.4), eliminare il surnatante ed aggiungere per ogni campione 100 ul di mix così composta: 73,5 ul di acqua HPLC, 10 ul di T4 Ligase buffer (10x), 10 ul di PEG-4000 (50%), 2,5 ul di Tween-20 (1%), 2 ul di adattatore CL53/73 (100uM) e 2 ul di T4 DNA ligasi (5 U/ul). Trasferire nel ThermoShaker secondo il seguente programma: 1 ora, 22°C, 1500 Hz. Utilizzando il rack magnetico eliminare il

surnatante e sospendere le biglie con 100 ul di buffer Stringency-Wash. Trasferire nel ThermoShaker secondo il seguente programma: 3 minuti, 45°C, 1500 Hz. Utilizzando il rack magnetico, eliminare il surnatante e aggiungere 200 ul di 0.1x BWT, eliminare nuovamente il surnatante ed aggiungere 50 ul di buffer EBT, incubare per 1 minuto a 95°C, trasferire in un rack magnetico e trasferire il surnatante contenente il prodotto della libreria a singolo filamento in tubini nuovi.

#### **4.6.6 Indicizzazione della libreria a singolo filamento**

Dopo quantificazione della libreria tramite PCR RealTime (Cf.4.5.5) si passa alla fase di indicizzazione della libreria genomica in maniera analoga a quanto fatto per la libreria a doppio filamento.

Per ogni campione aggiungere 20 ul di acqua HPLC, buffer Accuprime (con MgCl<sub>2</sub> e dNTPs) (10x), 1 ul di Accuprime Pfx DNA polimerasi (5U/ul), 10 ul di indice-P7, 10 ul di indice-P5 e 49 ul di libreria a singolo filamento, per un volume finale di 100 ul. Trasferire i campioni nel thermal cycle secondo il seguente programma: 2 min a 95°C, 20 secondi a 95°C, 30 secondi a 60°C, 1 minuto a 68°C, per 34 cicli. Successivamente si purifica, seguendo il manuale d'uso, con il kit MinElute PCR purification (QIAGEN) ed infine la quantificazione è effettuata con il Nanodrop (Thermo Scientific).

#### **4.6.7 PCR 1-cycle**

Effettuando l'indicizzazione della libreria con 34 cicli, la probabilità che si formino degli eteroduplici, ovvero delle sequenze di DNA appaiate male, è molto alta per cui per eliminarli viene effettuato un solo ciclo di PCR sul prodotto purificato dell'indicizzazione.

In un volume finale di 100 ul per ogni campione aggiungere 69 ul di acqua HPLC, 20 ul di buffer Herculanase II, 1 ul di dNTPs (2,5,mM), 4 ul di primer IS5 (10uM), 4 ul di primer IS6 (10uM), 1 ul di polimerasi Herculanase II (5U/ul). Trasferire nel thermalcycle secondo il seguente programma: 2 minuti a 95°C, 30 secondi a 95°C, 30 secondi a 60°C, 30 secondi a 72°C, 5 minuti a 72°C ed effettuare un solo ciclo. Purificare con il kit MinElute PCR purification (QIAGEN) seguendo il manuale d'uso.

A questo punto il campione è pronto per la cattura del DNA mitocondriale oppure con lo shotgun.

### **4.7 Cattura del DNA mitocondriale nelle librerie a doppio filamento**

Il DNA mitocondriale delle librerie a doppio filamento è stato catturato mediante il metodo Maricic (Maricic et al. 2010), descritto nella sezione 2 di questa tesi.

#### **4.7.1 Preparazione delle sonde**

Un estratto salivare viene amplificato con due Long Range PCR, usando il Kit Expan Long Range System, per ottenere l'intero genoma mitocondriale amplificato in due porzioni parzialmente

sovrapposte. Per questa amplificazione si usano due coppie di primers diversi con le seguenti sequenze:

**Prima coppia**

Primer Fwd: 5'-GGCTTTCTCAACTTTTAAAGGATA

Primer Rvs: 5'-TGTCCTGATCCAATCCAG

**Seconda coppia**

Primer Fwd: 5'-CCGTGCAAAGGTAGCATAATC

Primer Rvs: 5'-TTACTTTTATTTGGAGTTGCACCA

La Master mix per un campione prevede i seguenti reagenti: 10 µl di Buffer Expand Long Range System Roche 5X, 1,5 µl di DMSO al 100%, 1,5 µl di ciascun primer 10 µM, 2,5 µl di dNTP mix 10 µM, 3,7 U di Enzyme Long Range System Roche, 1 µl di BSA 20 mg/ml, 30,3 µl di H<sub>2</sub>O e 1 µl di estratto.

Secondo le istruzioni del produttore il profilo termico impostato per la reazione è:

- Step di attivazione: 92°C per 2 minuti

Per 40 cicli si ripetono i seguenti step termici:

- Denaturazione: 92°C per 10 secondi

- Annealing: 68°C per 15 secondi

- Estensione: 68°C per 9 minuti

Alla fine dei cicli

- Estensione finale: 68°C per 7 minuti

I prodotti delle Long Range PCR sono poi controllati mediante corsa su gel di agarosio e poi purificati con MinElute Purification Kit eluendo in 130 µl di TE low Buffer. I purificati vengono trasferiti in microtubi Covaris (appositi per l'ultrasonificazione) e sottoposti a sonicazione con il sistema Bioruptor per ottenere frammenti di DNA lunghi circa 300 bp, la lunghezza viene verificata nuovamente mediante elettroforesi su gel di agarosio. Il protocollo Maric prevede l'utilizzo di biglie magnetiche come supporto per le sonde, così è possibile separare dal resto della library il DNA ibridato semplicemente usando un magnete. Per questo le sonde vengono trattate in modo che possano aderire alla superficie delle biglie. Il primo step consiste nella riparazione delle estremità a singolo filamento dei frammenti sonicati (*Blunt-end Repair*): 76 µl di frammenti sonicati sono incubati a temperatura ambiente per 30 minuti in una master mix che è composta da Blunting Buffer 1X, una mix di dNTP 0,1 mM, Blunting Enzyme mix in un volume finale di 100 µl. Dopo la purificazione con MinElute Purification Kit (Qiagen), le sonde vengono biotinilate, grazie a una reazione catalizzata dalla Quick Ligase sono cioè legate a specifici adattatori biotinilati (Adaper Bio-T/B), che avviene durante l'incubazione 15 minuti a temperatura ambiente. Dopo l'incubazione i prodotti sono nuovamente purificati con MinElute Purification kit. Poiché è stato calcolato che solo 25 ng di DNA riescono a legarsi a 1 ng di biglie, i prodotti sono quantificati al Nanodrop

per conoscere la concentrazione delle sonde e poter massimizzare il legame tra sonde e biglie.

#### **4.7.2 Legame delle sonde alle biglie magnetiche**

Le biglie magnetiche sono rivestite con le sonde di mtDNA sfruttando l'interazione specifica biotina-streptavidina: le biglie magnetiche sono funzionalizzate con streptavidina in grado di riconoscere gli adattatori di biotina che abbiamo legato alle sonde nello step precedente. Poiché la streptavidina si stacca più facilmente dalle biglie quando è connessa a molecole di DNA è necessario preparare le biglie rivestite lo stesso giorno in cui si svolge l'esperimento di cattura. Si utilizza un volume di biglie pari a 5  $\mu$ l che come anticipato è in grado di legare 100 -125 ng di DNA esca, tuttavia il protocollo Maricic riporta che l'efficienza combinata di legame tra sonde e adattatori e tra adattatori e biglie è al massimo del 30%, pertanto è necessario disporre di 400-500 ng di sonde biotinilate per poter saturare completamente 5  $\mu$ l di biglie con il DNA esca. Quindi le biglie sono sottoposte a due lavaggi diversi: il primo con 200  $\mu$ l di buffer BWT 1X, composto da NaCl 2M, Tris-HCl 10 mM pH8, EDTA 1 mM pH8, Tween 20X 0,1% e H<sub>2</sub>O; il secondo con 200  $\mu$ l di TET. Mentre 50  $\mu$ l di sonde biotinilate sono denaturate in un ugual volume di buffer BWT 2X a 98°C per 1 minuto e velocemente messe in ghiaccio per evitare il riappaiamento fra i filamenti complementari. A questo punto si miscelano sonde e biglie incubandole a temperatura ambiente per 20 minuti su di un rotore così da facilitare l'interazione biotina-streptavidina e quindi il legame sonda-biglia. Usando un rec magnetico apposito si recuperano le biglie e si rimuove il supernatante contenente sonde non legate, le biglie sono poi sottoposte a due lavaggi con 200  $\mu$ l di Buffer BWT 1X riscaldato a 50°C e risospeso con 50  $\mu$ l di TET e conservate a 4°C. A questo punto le biglie sono rivestite di sonde e si può procedere con la cattura dell'mtDNA della libreria (Ibridazione).

#### **4.7.3 Ibridazione fra libraries e sonde**

Affinchè l'mtDNA delle libray indicizzate possa ibridarsi con le sonde denaturate i frammenti devono essere a singolo filamento, è quindi necessario utilizzare un sistema che eviti il riappaiamento dei filamenti complementari fra di loro e il *cross-linking* tra gli adattatori delle molecole *target* della libreria che vogliamo catturare e quelli di altre molecole non *target*. A questo scopo nel protocollo Maricic è prevista una fase preliminare all'ibridazione vera e propria nella quale le librerie sono incubate con *Blocking Oligo*, oligonucleotidi complementari agli adattatori Solexa Illumina P5 e P7. Inoltre la quantità di DNA da catturare deve essere pari a 2  $\mu$ l ed è possibile riunire più librerie in unico *pool*. Basandosi sulle quantificazioni effettuate precedentemente le librerie sono combinate in modo che il pool finale abbia un volume finale pari a 16,8  $\mu$ l, quello

necessario per la preparazione della *blocking-oligo hybridization mixture*. Quindi si prepara la Master mix con un set di *blocking-oligo* che per un campione contiene: 1 µl di BO4.part1.R (stock conc. 100 µM), 1 µl BO6P7.part2.R (stock conc. 100 µM), 1 µl BO8.P5.part1.R (stock conc. 100 µM), 1 µl BO11.splib.part2.R (stock conc. 100 µl), 5,2 µl Agilent blocking agent , 26µl Agilent hybridization buffer ( stock conc. 2X), per un volume totale di 35,2 µl. A questa aggiungiamo 16,8 µl di template (*pool* di library) ottenendo un volume finale di 52 µl e si lascia incubare per 3 minuti a 95°C e poi 30 minuti a 37°C. A questo punto si passa alla fase di cattura vera e propria aggiungendo i 52 µl di *hybridization mix* alle biglie magnetiche rivestite di sonde dopo aver rimosso il TET in cui erano conservate. La reazione avviene incubando questa miscela per 48 ore su un rotore a 12 rpm, per evitare la fuoriuscita di liquido durante l' incubazione avviene in tubini Safe-Lock sigillati.

#### **4.7.4 Eluizione delle librerie catturate**

Una volta che il DNA *target* si è ibridato alle sonde, si effettuano passaggi di lavaggio per eliminare la parte di libreria non ibridata. Prima si usa il rec magnetico per separare le biglie con l'mtDNA legato dal resto della soluzione, il surnatante viene scartato e successivamente le biglie sono state lavate tre volte con 200 µl di Buffer BWT 1X, eliminando ogni volta il surnatante utilizzando il magnete. Poi per due volte, le biglie sono risospese e incubate in 200 µl di Buffer HWT preriscaldato a 60°C per 2 minuti. Il Buffer HWT è composto da MgCl 2,5 mM, KCl 50 mM, Tris-HCl a pH 8 15 mM, 0,1% di Tween-20. A questo punto le biglie legate al DNA target sono risospese in 100 µl di TET e trasferite in nuove eppendorf da 1,5 ml per procedere con l' eluizione della libreria catturata. Per l' eluizione si utilizzano 50 µl di una soluzione di NaOH 125 mM capace di rompere i legami a ponte d' idrogeno formati fra filamenti di sonda e di libreria, rilasciando così in soluzione solo l'mtDNA catturato. Quindi sfruttando il magnete si recuperano le biglie e il surnatante contenente il DNA *target* viene trasferito in nuovi tubini. Per evitare che il DNA si degradi la soluzione di eluizione alcalina viene immediatamente neutralizzata aggiungendo una soluzione tampone preparata con 500 µl di PBI (kit MinElute) e 3,8 µl di acido acetico 20%. Per aumentare la resa della cattura il passaggio dell'eluizione viene ripetuto, incubando le biglie con altri 50 µl di NaOH facendo attenzione al viraggio del colore dell' indicatore di pH presente nel PBI, aggiungendo eventualmente altro acido acetico per mantenerlo basso. Gli eluiti sono purificati con MinElute Purification Kit (Quiagen) eluendo in 15 µl di TET e conservati a -20°C.

#### **4.7.5 Quantificazione e arricchimento del DNA catturato**

Per avere una prima informazione sul successo si esegue una Real Time PCR utilizzando i primers Bridge IS5 e IS6 complementari alla sequenza

degli indexing-primer esterna a quella degli indici, con lo stesso kit e protocollo precedentemente descritti ma diluendo i catturati 1:10, il processo di cattura riduce drasticamente la quantità di DNA. Anche in questo caso i valori di concentrazione ottenuti con la qPCR servono per calcolare i cicli della successiva *Accuprime PCR*, eseguita con lo stesso protocollo già descritto, al fine di amplificare i catturati. A questo punto si procede con il sequenziamento nella piattaforma Illumina.

## **4.8 Cattura del DNA mitocondriale nelle librerie a singolo filamento**

### **4.8.1 Preparazione delle sonde.**

Sono state generate 5506 sonde utilizzando il DNA mitocondriale moderno come stampo. Dopo rimozione di eventuali sonde ripetute, sono stampate innumerevoli volte generando un array contenente circa 973,304 sonde. L'array così costruito viene convertito in una libreria di sonde secondo il procedimento seguente: dopo aggiunta di 500  $\mu$ l di Elution Solution (125mM NaOH pH8,5, 0,05% Tween 20), ogni array è assemblato in una gasket slide posta in una camera di ibridazione e messo in un rotore per 6 ore a 12 rpm a temperatura ambiente. L'eluato è poi recuperato con una siringa, neutralizzato aggiungendo 19  $\mu$ l di acido acetico al 20% e purificato con il kit QIAquick Nucleotide Removal (QIAGEN). Le sonde così purificate sono eluite in 20  $\mu$ l di EB buffer (10 mM Tris-Cl pH8,5, Qiagen) e confermatone il recupero tramite denaturazione PAGE che permette la separazione delle molecole di DNA su gel di acrilamide grazie all'azione dell'SDS.

#### **-Legame con il primo adattatore**

Il primo adattatore è stato aggiunto alla libreria di sonde secondo la seguente reazione in un volume finale di 50  $\mu$ l: 10 $\mu$ l sonde purificate, 3  $\mu$ l Bst polymerase, 1X Thermopool buffer, 250 mM di ogni dNTPs, 1  $\mu$ l di primer APL2, amplificazione secondo il profilo termico 1 min 50°C, 5 min 15°C, 5 min 20°C, 5 min 25°C, 5 min 30°C e 10 min a 37°C. La reazione è stata poi purificata con Minelute PCR purification kit.

#### **- Blunt and Repair**

Dopo l'attacco del primo adattatore si procede con la reazione di Blunt and Repair per eliminare al 3' le differenze tra le sequenze del filamento di DNA lasciate dalla Bst polymerase. In un volume finale di 40  $\mu$ l vengono aggiunti: 20  $\mu$ l di prodotto purificato della reazione precedente (legame con 1 adattatore), 0,4  $\mu$ l di T4 DNA polymerase, 0,4  $\mu$ l di Frammenti di Klenow, 1x buffer Tango e 100 $\mu$ M di ogni dNTPs. La reazione così composta viene incubata 15 min a 25°C, successivamente urificata con MinElute PCR purification kit (QIAGEN) ed eluita in 20 $\mu$ l di EB buffer.

#### **-Legame con il secondo adattatore**

Il secondo adattatore che viene aggiunto alla libreria di sonde è a doppio filamento e viene assemblato in una reazione composta da 7  $\mu$ l di acqua, 1  $\mu$ l di T4 ligase buffer, 1  $\mu$ l di 100 $\mu$ l di ogni dNTPs, 1  $\mu$ l di 100  $\mu$ M primer

APL6, secondo il profilo termico: 10 sec a 95°C per poi raggiungere lentamente i 14°C scendendo di 0,1°C al secondo. Successivamente vengono aggiunti 3 ul di T4 DNA buffer, 2 ul di acqua, 4 ul di PEG 4000 al 50%, il purificato del Blunt and Repair, 1 ul di T4 DNA polymerase ed incubata per 30 min a temperatura ambiente purificando infine con il kit MinElute PCR purification.

Per rimuovere i dimeri degli adattatori le biglie (20ul di MyOne C1 Streptavidin) sono lavate due volte con BWT + SDS buffer (1M NaCl 10mM Tris-HCl pH 8, 1mM EDTA, 0,05% Tween-20, 0,5% SDS). Le biglie sono poi sospese in 180 ul di BWT+SDS e viene aggiunto tutto l'eluito (20ul) e messo nel rotore per 20 minuti a temperatura ambiente. Successivamente le biglie sono lavate due volte con il buffer 0,1X BWT e sospese in 25 ul di TT buffer, incubate poi per 10 minuti a 95°C, in questo modo vengono rilasciati i filamenti biotinilati che rappresentano le sonde pronte per catturare il DNA mitocondriale presente nelle librerie genomiche. Il numero delle molecole di sonde sono determinate tramite qPCR usando il set di primers APL5 e APL6. Successivamente le sonde sono amplificate in un volume finale secondo la seguente reazione: 10 ul di sonde 1ulHerculase II Fusion DNA polymerase 1X Herculase II buffer, 250uM di ogni dNTPs, 400 mM dei primers APL5 e APL6 secondo il profilo termico 2 min. a 95°C, 20 sec. a 95°C, 60 sec. 60°C, 30 sec a 72°C, 5 min. a 72°C. Segue purificazione con MinElute PCR purification kit ed eluizione finale in 20ul di EB buffer (Qiagen). Successivamente si caricano su gel di agarosio low melting 7 ul di ogni prodotto di amplificazione e le bande di circa 94pb vengono rimosse per eliminare gli smear. Il DNA è isolato dal gel con il kit MinElute gel extraction (Qiagen) ed eluito in 30ul di EB (Qiagen). Il successo dell'estrazione è valutato tramite qPCR usando tra 8 e 19 ul di banda tagliata e amplificando in una reazione dal volume finale di 100ul composta da: 1ul Herculase II buffer 250 uM per ogni dNTPs 400 mM dei primers APL5 e APL4, secondo il profilo termico 2 min a 95°C, 20 sec a 95°C, 30 sec a 60°C, 30 sec a 72°C, 5 min a 72°C per 8 cicli. Il prodotto è purificato con Minelute PCR purification kit, eluito in 40ul di EB e quantificato con il Nanodrop, la concentrazione varia tra circa 69 e 97 ng(ul. Le sonde a singolo filamento sono generate in una amplificazione "SINGLE-PRIMED LINEAR" usando un unico primer biotinilato (APL2). Questa reazione prevede una nuova amplificazione di 200ul di sonde con 2ul Herculase II buffer 250uM per ogni dNTPs 400 nM ddi primer APL2 secondo il profilo termico 2 min 95°C 20 sec 95°C 20 sec 60°C 20 sec 72°C 20 sec 72°C per 20 cicli. I prodotti dell'amplificazione sono uniti insieme a formare un unico pool e mischiati con un volume doppio al 38% di PEG-SPRI (Solid Phase Reversible Immobilization) (circa 20ml). Quest'ultimo è un sistema di purificazione basato sull'utilizzo di biglie magnetiche nel quale il legame con il DNA è rese reversibile dalla presenza di PEG

(Polyethylene glycol). Le sonde sono poi eluite in 150ul di TE e quantificate con il Nanodrop conservandole poi a -20°C.

#### **4.8.2 Ibridazione fra librerie e sonde**

Ogni cattura è formata da 6ul di libreria a singolo filamento (circa 2ng), 5,25 ul di acqua HPLC, 2,5ul 1 mg/ml di Cot-1-DNA per saturare le sequenze ripetitive del genoma umano, 0,25ul 10mg/ml di DNA estratto da sperma di salmone che ha funzione bloccante, 0,5ul 500uM BO4, 0,5 ul 500uM BO10. Ogni cattura è incubata per 5min a 95°C e 5 min a 65°C mantenute poi a temperatura ambiente. Le sonde da aggiungere sono formate diluendo in acqua 300ng di sonde a singolo filamento ottenendo un volume finale di 4ul. La reazione di ibridazione è posizionata nei 96 pozzetti di una piastra e successivamente vengono aggiunti 20 ul di Buffer di Ibridazione (1,7ul di Hi-RPM buffer (Agilent) e 300 ul 50X Denhardt's Solution) ed infine le sonde. La piastra è vortexata e incubata a 62°C per 2 giorni.

#### **4.8.3 Preparazione delle biglie:**

30ul di biglie magnetiche MyOne T1 sono lavate con 150ul di Wash buffer (1X SSC, 0,01% SDS) per 15 min a temperatura ambiente, successivamente sono lavate tre volte con 120 ul di HWT buffer (1X Amplitaq gold buffer senza MgCl<sub>2</sub>, 0,02% Tween-20) per 10 min a 60°C e nuovamente lavate con 150ul di Wash buffer. Prima di ogni passaggio il liquido percolato viene rimosso. Dopo l'ultimo lavaggio le biglie sono eluite in 19ul di Melt Solution (125ul NaOH, 0,05% Tween-20) ed incubate 15 minuti a temperatura ambiente. Successivamente il supernatante viene trasferito in una nuova piastra a cui vengono aggiunti 0,7ul di acido acetico al 20% e 190ul di PN buffer (Qiagen). Ogni cattura viene poi purificata con MinElute Nucleotide Removal kit. Il DNA è poi eluito in 30ul di buffer TT. Per verificare il successo della cattura viene effettuata una qPCR utilizzando 1ul di catturato. I rimanenti 29ul vengono amplificati in un volume finale di 100ul in una reazione contenente: 1xHerculase II buffer 250uM di ogni dNTPs, 400nM dei primers genomici R1 e multiplex R2, secondo il seguente profilo termico: 2min 95°C 30sec 95°C , 30 sec 60°C, 30 sec 72°C 5 min 72°C per 28 cicli. Gli amplificati sono purificati usando un rapporto di 1:1 di PEG-SPRI al 38%, eluendo poi in 15ul di TE determinandone la concentrazione con il Nanodrop. 3ul di catturato (circa 500ng) derivante dalla prima ibridazione sono usati come templatato per una seconda ibridazione seguendo la procedura descritta finora eccetto per una diminuzione del tempo di ibridazione che da due giorni passa ad un giorno. Segue poi, come già descritto, quantificazione tramite qPCR, amplificazione per 23 cicli, purificazione con biglie magnetiche ed ultima quantificazione con Nanodrop. Dopo la purificazione con MinElute PCR purification kit il risultato della cattura viene analizzato e visualizzato con Agilent.

## 4.9 Sequenziamento

Le librerie catturate sia a doppio che a singolo filamento e una parte della libreria a singolo filamento non catturata sono state sequenziate su piattaforma Illumina MiSeq utilizzando il kit V3 impostando 2x76 cicli di sequenziamento. Le sequenze ottenute sono state sottoposte al processo di *de-multiplexing* attraverso il quale le sequenze di ogni campione sono separate in base mediante ai diversi indici assegnati.

## 4.10 Analisi Bioinformatica

### 4.10.1 Analisi bioinformatica dei risultati del sequenziamento delle librerie a doppio filamento

#### Adapter trimming e merging delle reads

Prima di iniziare l'effettiva analisi dei dati ottenuti, la qualità del sequenziamento è controllata utilizzando il programma FASTQC. Questo strumento fornisce diversi moduli di analisi per individuare varie caratteristiche delle sequenze, come il loro numero totale, la lunghezza, la qualità del sequenziamento per base e per sequenza ed il contenuto in GC. Valutati i report le sequenze vengono preparate per il passaggio successivo che è rappresentato dal *mapping assembly*, utilizzando un pacchetto di script denominati fastqProcessing, programmati in linguaggio Python (Kircher 2012). Partendo dall'assunto che il DNA antico è caratterizzato da frammenti relativamente brevi e che le reads contengano parte o l'intera sequenza degli adattatori a valle dell'inserto, si applicano dei sistemi bioinformatici in grado di effettuare l'*adapter trimming*, eliminando quindi la sequenza degli adattatori. Un punto molto importante dell'analisi bioinformatica è il *merging*, ovvero l'unione delle *reads* R1 e R2, derivanti dal sequenziamento in paired end, con minimo 10 pb di sovrapposizione (*overlap*). In questo modo viene ricostruita la sequenza intera che verrà utilizzata negli step successivi di analisi. Inoltre le reads sono appositamente filtrate in modo da eliminare le sequenze con un alto contenuto in GC e quelle con più di 5 nucleotidi con un valore di qualità di sequenziamento (PHRED score) minore di 15.

#### Mapping assembly dell' mtDNA

A questo punto si effettua il mappaggio delle *reads merged* filtrate sul genoma mitocondriale utilizzando il Mapping Iterative Assembler (MIA), composto da una serie di script in grado di allineare tutte le reads relative a una libreria con una sequenza di riferimento creando una sequenza consenso. Questo programma è appositamente ideato per il DNA antico, infatti per l'allineamento e la creazione della sequenza consenso, utilizza una matrice di sostituzione posizione-specifica che tiene conto dei possibili danni presenti sull'aDNA. Inoltre dato che per l'aDNA il flusso di lavoro prevede

vari passaggi di amplificazione per aumentare la quantità di DNA endogeno, il primo *script* di 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 *missmatch*. Questo è importante perché un alto numero di *reads* identiche può portare a stime errate della frequenza di polimorfismi. L'*output* dello script sono uno o più files (con estensione .maln) *unique*, contenenti le reads copie della stessa sequenza collassate e da uno o più files (con estensione .maln) *not-unique*, nei quali le *reads* sono state mappate mantenendole separate. I file *unique* e *not-unique* sono poi usati come input per gli script successivi fino per creare la sequenza consenso mappando sulla sequenza di riferimento CRS (Anderson et al. 1981; Andrews et al. 1999). Il rapporto fra reads *unique* e *not-unique* permette di definire un *cluster factor* che da una stima della profondità di sequenziamento, mentre attraverso un altro script siamo in grado di calcolare la percentuale di copertura dell'mtDNA e della profondità del sequenziamento per ogni singola posizione.

### **Burrows-Wheeler Aligner**

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

### **MapDamage**

Il pattern di misincorporazione viene determinato, misurando lo stato di degradazione del DNA attraverso il programma MapDamage 2.0, un framework scritto in Python in grado di tracciare e quantificare i pattern di danneggiamento all' interno dei dati di sequenza di aDNA generati da piattaforme NGS (Ginolhac et al. 2011; Jönsson et al. 2013). Questo modello esegue una stima secondo un metodo bayesiano di vari parametri di danno al DNA. Uno dei parametri considerati è la transizione da citosina a timina (C→T), causata dalla deaminazione *post-mortem* della citosina che avviene maggiormente a carico di molecole a singolo filamento. Dato che i fenomeni degradativi portano alla formazione di frammenti con sporgenze

alle estremità è chiaro che la frequenza degli eventi di transizione C→T è maggiore alle estremità di una molecola piuttosto che in una porzione interna a doppio filamento. Dato il procedimento di preparazione delle librerie questa caratteristica risulta come un eccesso di misincorporazioni C→T al 5' delle sequenze e in un eccesso di misincorporazioni G→A al 3'. Il programma è impostato per analizzare il tasso dei vari tipi di sostituzione oltre che delezioni e inserzioni, fino al nucleotide 70. Il rapporto fra numero di transizioni C→T e numero totale di C all' estremità 5' e tra numero di G→A e numero totale di G al 3' permette di calcolare la frequenza di ciascuno dei due tipi di misincorporazione solitamente fino alla posizione 10 (Briggs et al 2007).

#### **4.10.2 Analisi bioinformatica dei sequenziamenti delle librerie a singolo filamento**

L'analisi bioinformatica è stata condotta utilizzando i software Bustard (Illumina) o freeIbis per assegnare le basi durante il sequenziamento. Le sequenze che non combaciano perfettamente con gli indici che sono stati inseriti vengono scartate. Per effettuare il merging, ovvero la ricostruzione completa delle sequenze, sono stati usati appositi scripts elaborati per far fronte alle necessità del DNA altamente frammentato come quello che caratterizza i reperti scheletrici antichi. Le sequenze che hanno effettuato il merging e risultano essere maggiori di 30 pb sono allineate con la sequenza di riferimento che in questo lavoro risulta essere la rCRS (revised Cambridge reference sequences, NC\_012920) usando il software BWA impostato con i seguenti parametri, '-n5' per permettere 5 mismatches e '-l16500' per considerare la lunghezza della sequenza. Le sequenze con identiche coordinate di inizio e fine sono collasate in una singola sequenza. Per diminuire la presenza dei contaminanti umani e per eliminare allineamenti spuri sono stati usati una serie di filtri:

- esclusione delle sequenze con un valore di deaminazione inferiore al 5%
- mappaggio delle sequenze con BWA permettendo 5 mismatches e una inserzione o delezione per prevenire la perdita di sequenze con alcuni cambi C-T
- rimozione delle sequenze che mostrano più di due differenze con la sequenza di riferimento che non possono essere spiegate con la deaminazione della citosina
- le differenze con la sequenza di riferimento sono limitate a cinque considerando anche inserzioni o delezioni
- le sequenze con un valore di qualità del mappaggio inferiore a 30 sono rimosse.

#### **Ricostruzione della sequenza consenso**

Per ovviare al fatto che tutte le librerie sono contaminate con DNA moderno, vengono usate le sostituzioni C/T per arricchire le sequenze endogene prima

di ricostruire la sequenza consenso. Questo approccio per identificare le sequenze che presentano la deaminazione-indotta è determinato effettuando la comparazione con la sequenza di riferimento per il genoma mitocondriale. I contaminanti umani potranno occasionalmente mostrare delle reali mutazioni C/T rispetto alla sequenza di riferimento dovuto alla divergenza tra le sequenze. Per ovviare a quest'ultima problematica è stato sviluppato un altro approccio nel quale vengono isolate le sequenze con una T terminale (o con una T nelle prime tre posizioni o nelle ultime tre) se almeno l'80% di tutte le sequenze coprono la rispettiva posizione nel DNA mostrano una C. Questa procedura considera il fatto che il cambio C/T non è indicativo di una deaminazione se condiviso con molte altre sequenze indipendentemente dalla sequenza di riferimento. Per ridurre gli effetti del danno-indotto dall'arricchimento con C/T durante la ricostruzione della sequenza consenso le Ts vengono convertite in Ns nelle prime ed ultime tre posizioni di ogni sequenza. In queste posizioni le citosine sono convertite in uracile con una frequenza maggiore del 10%. Le T sono convertite in N se almeno un'altra sequenzamostra una C nella stessa posizione. Lo script Mpileup, appartenente al pacchetto di SAMtools, è stato utilizzato per convertire il file con estensione BAM in un formato tabulare basato sulla posizione. Le tabelle così generate sono utilizzate per determinare la copertura di ogni posizione, la base consenso determinata come la base maggiormente rappresentata e la percentuale di sequenze che supportano la base consenso. Dato che BWA non tiene in considerazione la circolarità della molecola del DNA mitocondriale, mappaggio, filtraggio e costruzione della sequenza consenso sono ripetuti usando una sequenza di riferimento modificata nella quale 1Kb di sequenze sono rimosse dall'inizio alla fine. In questo modo si ottiene la stessa misurazione anche per la prima e l'ultima base del DNA mitocondriale.

### **Costruzione della sequenza consenso con differenti parametri di filtraggio**

Un'accurata ricostruzione della sequenza consenso è complicata a causa dell'elevatissima quantità di contaminanti moderni e dalla lunghezza ridotta delle sequenze endogene. I frammenti corti di DNA sono meno efficientemente arricchiti durante la cattura per ibridazione, inoltre gli errori durante l'assemblaggio possono ridurre la probabilità di identificare le sequenze endogene se differiscono dalla sequenza di riferimento. Sia la contaminazioni che gli errori di mappaggio, se non vengono rimossi, potrebbero rendere la sequenza consenso del campione più simile di quanto lo sia in realtà ad una sequenza mitocondriale moderna. Per cercare di diminuire queste problematiche vengono introdotti degli ulteriori metodi di filtraggio delle sequenze di DNA:

- La sequenza consenso viene ricostruita previo utilizzo di filtri per cui la posizione può essere determinata in modo accurato e preciso.

Questo primo filtro si basa sull'osservazione della presenza della sostituzione C to T nella prima ed ultima base di ogni sequenza, in questo modo aumenta il potere discriminante tra le sequenze endogene ed i contaminanti. Viene inoltre usato un minimo di 5 coverage e la sequenza consenso deve essere almeno l'80% della sequenza consenso.

- Il secondo filtro applicato per la costruzione della sequenza consenso è meno stringente del precedente, infatti viene usato un minimo di 3 fold e la sequenza consenso deve essere almeno il 65% rispetto alla sequenza di riferimento. Inoltre viene osservata la presenza della mutazione C to T nelle prime ed ultime tre posizioni di ogni reads, aumentando così il numero di sequenze utili per le successive analisi bioinformatiche.

## 5. RISULTATI

### 5.1. Pulizia e Polverizzazione

La pulizia e polverizzazione di "Paus" (Fig 16) sono state interamente svolte presso il Laboratorio di Antropologia molecolare, Università degli Studi di Firenze, seguendo tutte le norme per evitare le contaminazioni ambientali e da DNA moderno. Il foro per la polverizzazione è stato praticato nello spessore della dipole cranica (Fig.17) nella zona del temporale destro, in accordo con il prof. G. Manzi, antropologo fisico dell'Università la Sapienza di Roma per non interferire con le eventuali analisi successive. Sono stati prelevati globalmente circa 200mg di polvere.



**Figura16:**Osso frontale di Paus



**Figura 17:** Particolare durante la polverizzazione

## 5.2 Libreria a doppio filamento

Sono state preparate due librerie a doppio filamento (Meyer and Kircher 2010), partendo dallo stesso estratto utilizzando per entrambe 20  $\mu\text{l}$  di DNA ed è stato preparato un controllo negativo della libreria (Klibreria). Una volta effettuati i passaggi di Blunt-end Repair, Adapter Ligation e Fill-In (cf. “Materiali e Metodi”), è stata effettuata la quantificazione mediante Real-Time PCR (RT-PCR). I risultati ottenuti (**Tabella 1**) mostrano che per entrambe le librerie, ma soprattutto per la prima, sono state raggiunte elevate quantità di DNA (Tabella 1 colonna 2) così come sono stati prodotti buoni valori dopo aver effettuato l’indicizzazione delle molecole (Tabella 1 colonna 3). Dal numero di molecole è possibile calcolare la quantità di DNA totale necessaria a determinare in quante aliquote dividere ogni campione in modo che alla fine dell’inserimento degli indici, si raggiunga una concentrazione di circa  $10^{13}$  (plateau di PCR) senza però superarla. Una volta calcolato il numero di cicli necessari si procede con l’indicizzazione della libreria (*indexing PCR*). Per ogni campione è stata scelta una coppia di indici diversa tra quelli disponibili (Tab.2).

1 Campione	2 N° copie totali indexing pre-	3 N° copie totali post- indexing (mol/ $\mu\text{L}$ )	4 Fattore di Amplificazione
1° Libreria	3,20E+009	1,12E+010	3,5
2° Libreria	1,08E07	1,06E09	98,14

**Tabella 1:** Quantificazione tramite RT-PCR delle molecole di DNA al termine della libreria genomica (N° copie totali pre-indexing) e dopo l’aggiunta degli indici (N° copie post-indexing) e calcolo del fattore di amplificazione

Il prodotto della indicizzazione è stato quantificato mediante RT-PCR, utilizzando primers che si appaiano alle sequenze esterne agli indici. Il risultato ottenuto rappresenta il numero di molecole effettivamente indicizzate con il quale è stata calcolata l'efficienza dell'indexing (Tabella 1 colonna4) secondo la formula:

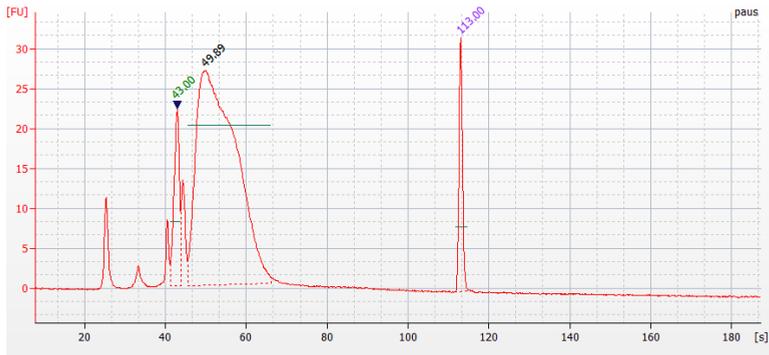
**Fattore di Amplificazione= Numero copie Post-indexing/Numero copie Pre-Indexing**

Il fattore di amplificazione risultante per le due librerie è 3,50 e 98,14. In generale ci aspettiamo che il numero di molecole di DNA aumenti dopo l'*Indexing PCR* dato che quest'ultima determina un'amplificazione del DNA, ed in effetti è ciò che abbiamo ottenuto (Tabella 1), anche se la prima libreria ha avuto un'efficienza di amplificazione minore probabilmente legata all'elevata quantità di materiale iniziale. Per cercare di ottenere la più alta quantità di DNA possibile, le librerie indicizzate sono state ulteriormente amplificate mediante una reazione di PCR che vede l'impiego dell'enzima Accuprime Pfx DNA polymerase. Il numero di cicli da effettuare sono scelti in base alla quantità di DNA ottenuta nell'indicizzazione e cercando sempre di mantenere la concentrazione di DNA sotto il *plateau* della reazione (1E+13). In particolare sono stati effettuati 11 cicli per la prima libreria e 13 per la seconda.

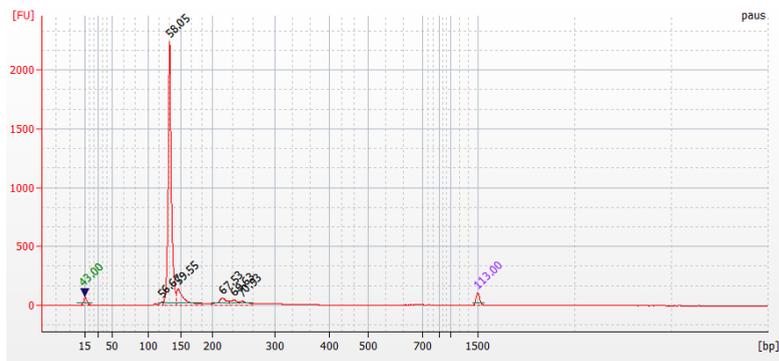
Campioni	P7 primer ID	P7 sequence	P5 primer ID	P5 sequence
1°Libreria	Sol_iPCR-MPI-3	ATCAG	P5_iPCR-LP-9	ACCAACT
2°Libreria	Sol_iPCR-MPI-10	AAGCGA	P5_iPCR-LP-9	ACCAACT

**Tabella 2:** Indici utilizzati

Il prodotto di questa amplificazione viene analizzato con Agilent 2100. Per la prima libreria è ben visibile il picco formato dalle molecole di DNA della lunghezza media di 72 pb con una concentrazione di 53,33 ng/ul (Fig. 18). Mentre l'Agilent per la seconda libreria (Fig. 19) presenta scarsi picchi di DNA con lunghezza dei frammenti compresa tra 211pb e 247pb con una concentrazione media di 18,5 ng/ul ma con in corrispondenza del frammento di 133 pb la concentrazione risulta essere 88 ng/ul.



**Figura 18:** Risultato per la prima libreria post-amplificazione di Paus



**Figura 19:** Risultato per la seconda libreria di Paus post-amplificazione

Il profilo delle due librerie non presenta sequenze lunghe, tra le 500 e le 800 pb, che potrebbero rappresentare contaminanti di DNA moderno, per i quali la lunghezza dei frammenti attesa è maggiore (Krause et al.2010) oppure artefatti dovuti per esempio alla produzione di molecole ricombinanti generate dal legame degli adattatori alle estremità dei frammenti durante le amplificazioni tramite PCR. Per ovviare a quest'ultimo problema può essere effettuato un singolo ciclo di amplificazione con Accuprime Pfx PCR polymerase. Nei profili ottenuti, sono visibili anche picchi di lunghezze fino a 135bp che indicano la formazione di dimeri di adattatori, che nella seconda libreria mostrano una concentrazione particolarmente alta. La presenza di dimeri di adattatori può essere dovuta a una non perfetta purificazione durante il flusso di lavoro, ma è un problema che, in generale, affligge tutti i protocolli di preparazione delle librerie (Gansauge e Meyer 2013).

### 5.2.1 cattura del DNA mitocondriale

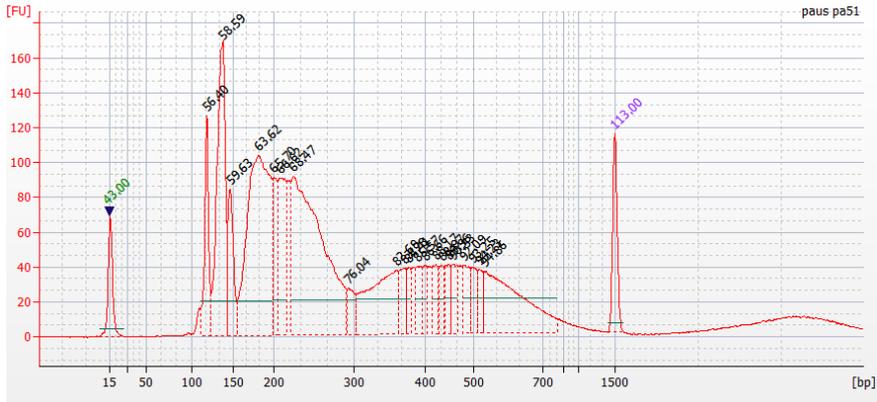
La cattura del DNA mitocondriale è stata eseguita secondo il protocollo messo a punto da Maricic e collaboratori (Maricic et al. 2010) che prevede l'utilizzo di frammenti amplificati di mtDNA umano utilizzati come sonde. Le librerie sono state riunite a coppie con una concentrazione totale di 2000 ng, prima di incubarli con la mix di Ibridazione (cf. "Materiali e Metodi"). In seguito il DNA catturato è stato quantificato mediante RT-PCR utilizzando i Primer Bridge IS5 e IS6 complementari alle sequenze esterne agli indici. Come evidenziato dai risultati mostrati in Tabella 3 è evidente una netta diminuzione della

1 Campione	2 N° Copie Post-Indexing (mol/μL)	3 N° Copie Post-Cattura (mol/μL)
Paus 1libreris	3.55E09	8.58E05
Paus 2°Libreria	1,06E09	1,18E06

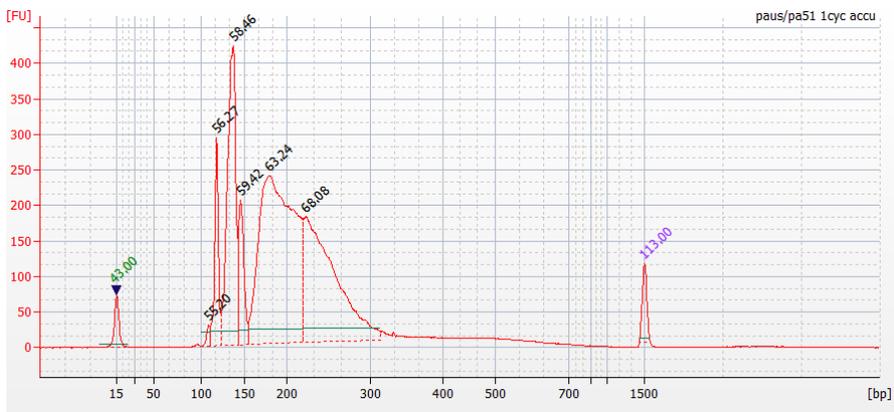
**Tabella 3:** In colonna 3 è mostrata la quantità delle molecole di DNA post catturato determinata tramite la RT-PCR

concentrazione del DNA dei campioni catturati (Tab. 3 ) rispetto a quello delle librerie indicizzate (Tab.3). Considerando che le sonde ibridano solo con l'mtDNA, la caduta di concentrazione notata è in accordo con l'atteso. Per aumentare la concentrazione dell'mtDNA catturato, al fine di migliorare il processo di sequenziamento, è stata eseguita nuovamente una PCR utilizzando sempre gli stessi primers Bridge IS5 e IS6, e modulando i cicli, basandosi sui risultati ottenuti dalla RT-PCR (Tabella 3), per fare in modo che il numero di molecole non sia maggiore di  $10^{13}$ .

Una volta purificate, le librerie catturate sono state quantificate con Agilent BioAnalyzer 2100. I risultati per la prima libreria catturata (Figura 20) mostrano un insieme di picchi principali di DNA che vanno da 146 a 292 pb con una concentrazione massima di 21ng/ul, compatibile con la presenza di DNA a doppio filamento. Le altre sequenze di lunghezza comprese tra 300pb e 500 pb, sono quasi sicuramente eteroduplici, per cui è stato necessario effettuare un unico ciclo di amplificazione con Accuprime Pfx DNA polymerase (1-cycle PCR) per cercare di diminuire la concentrazione dei frammenti lunghi. Come mostrato in Figura 21 il profilo appare pulito, con un range di frammenti di DNA compreso tra 109 e 223 pb con una concentrazione massima di 70 ng/ul in corrispondenza di 180 pb.

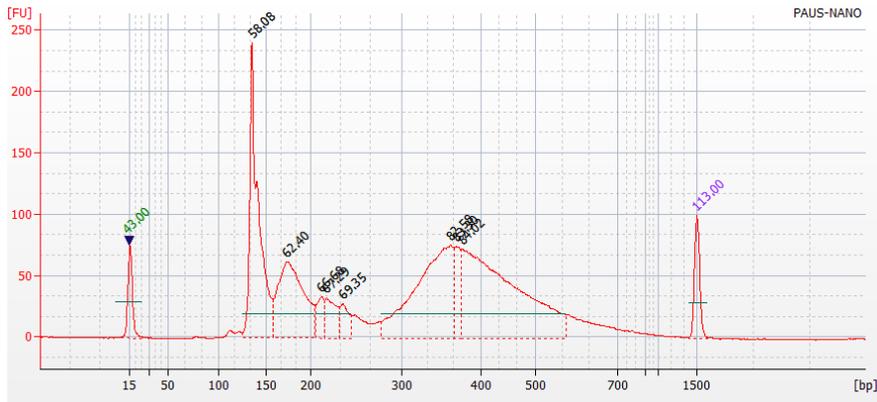


**Figura 20:** Risultati visualizzati all' Agilent 2100 per la cattura della prima libreria di Paus.



**Figura 21:** Risultati visualizzati con Agilent 2100 per la cattura della prima libreria dopo 1-cycle PCR

Riguardo alla cattura della seconda libreria di Paus i risultati ottenuti (Fig.22) mostrano la presenza di picchi tra 173 pb e 376 pb con una concentrazione massima di 15,24 ng/ul. In questo caso non è stato reputato necessario effettuare un singolo ciclo di amplificazione perché tutte le sequenze erano comunque all'intero di 300pb.



**Figura 22:** Risultati visualizzati all'Agilent 2100 per la cattura della seconda libreria di Paus

### 5.2.2 Analisi Bioinformatica

Le sequenze ottenute sono state sottoposte al processo di *de-multiplexing* e separate così in base ai diversi indici assegnati. Le sequenze sono state poi analizzate con scripts appositamente elaborati per l'analisi del DNA degradato. Le sequenze sono state analizzate in base alla qualità di sequenziamento, e poi manipolate utilizzando *fastqprocessing*, un pacchetto di 6 script appositi grazie al quale le reads forward (R1) e reverse (R2) subiscono l'*adapter trimming* e il *merging*; dato che abbiamo reads di 75 bp, per il merging è scelto un range di sovrapposizione di 10bp, così da selezionare solo *reads* in grado di dare origine a frammenti minori di 140 pb. Come risulta dalla Tabella 4 il numero delle sequenze ottenute non è molto elevato (Tab.) e dopo il merging si assiste ad una ulteriore diminuzione (Tab.4). La percentuale delle sequenze che hanno effettuato il merging risulta essere dell'80% per la prima libreria e del 15% per la seconda (Tab.4). Vengono inoltre applicati dei filtri che selezionano le sequenze in base alla qualità del sequenziamento (Tab.4 colonna 6) restringendo ancor di più la quantità utile di sequenze da analizzare.

1 Campione	2 Original Numero reads	3 Merged reads	4 Reads totali separate	5 % Merged	6 Filtri di qualità (sequenze totali)
1°Libreria	701022	284536	71070	80,01	318826
2°Libreria	20784	2151	11526	15,72	6449

**Tabella 4:** Risultati per il merging e quality filter

Successivamente le sole reads merged sono state sottoposte al *mapping assembly*, sono cioè mappate sulla sequenza di riferimento per *Homo*

*Sapiens* (rCRS) utilizzando il software Mapping Iterative Assembler (MIA), un pacchetto di 7 script specifico per il DNA degradato, in grado di restituire una sequenza consenso, tenendo conto dei possibili danni presenti nell'aDNA (cf Materiali e Metodi). Il primo di questi script ha collassato tutte le *reads* allineate con la stessa direzione e le stesse coordinate di inizio e fine in una sequenza unica dato che rappresentano cloni delle stesse molecole originali, prodotti durante la generazione dei cluster di frammenti dalla piattaforma Illumina.

1 Campione	2 Not unique	3 Unique reads	4 Cluster factor	5 Average coverage	6 Average length
1°LIBRERIA	1480	395	3,74	2,093	87,74
2°LIBERIA	389	336	1,15	1,777	87,58

**Tabella 5:** Risultati del software MIA

Inoltre ha allineato tutti i frammenti separati e quindi non unici (Tab.5 colonna 2) restituendo quindi per ogni campione anche uno o più *files* in base alla difficoltà con la quale effettua il mappaggio. Le sequenze uniche (Tab.5 colonna 3) risultano molto inferiori rispetto alle non uniche, mostrando che i dati ottenuti per la prima libreria sono più scarsi rispetto a quelli della seconda nella quale il numero di sequenze quasi si equivale. Il rapporto fra sequenze non uniche (not unique) e uniche (unique) produce il “Cluster Factor”(Tab 5, colonna 4). Questo parametro indica quante copie di ogni frammento sono state in media sequenziate e fornisce una stima del grado di profondità di sequenziamento, sono considerati ottimali valori compresi fra 2 e 10, valori minori sono indice di un sequenziamento poco profondo, mentre valori maggiori di un sequenziamento ridondante Per la prima libreria questo valore raggiunge circa 3 (Tab 5) indicando quindi un sequenziamento leggermente ridondante, mentre per la seconda libreria questo valore è pienamente nel range ottimale (Tab 5). Un altro valore importante che il pacchetto MIA restituisce è la copertura media, “Average Coverage”, calcolato in base a quante sequenze uniche coprono ogni singola base del DNA mitocondriale (Tabella 6). Sfortunatamente le due librerie di Paus presentano valori non particolarmente alti. Da sottolineare che questo è un dato aspettato visto che il numero di sequenze uniche così come il numero di merged è risultato molto basso. La copertura del genoma mitocondriale è anche determinata calcolando la percentuale dei nucleotidi che sono coperti dai frammenti unici da 1 a 5 volte (1fold, 2fold, 3fold, 5fold).

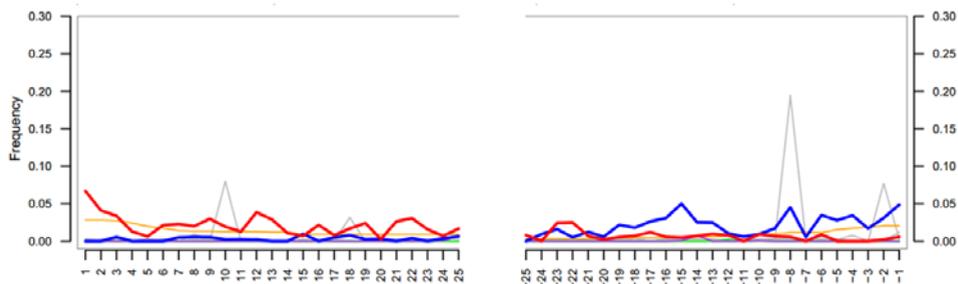
Campione	1-fold%	2-fold%	3-fold%	4-fold%	5-fold%
1°LIBRERIA	84,63	59,48	34,92	16,60	8,00
2°LIBRERIA	80,79	50,88	26,48	11,35	4,47

**Tabella 6:** Risultati di Fold Coverage per le librerie di Paus

I risultati per entrambe le librerie (Tab.6) decrescono drammaticamente dall'uno fold verso il cinque fold raggiungendo percentuali del 8% e 4,47%. La sequenza ottenuta con quest'ultima copertura viene utilizzata per le analisi successive per cui percentuali così basse indicano che non è possibile la ricostruzione del DNA mitocondriale.

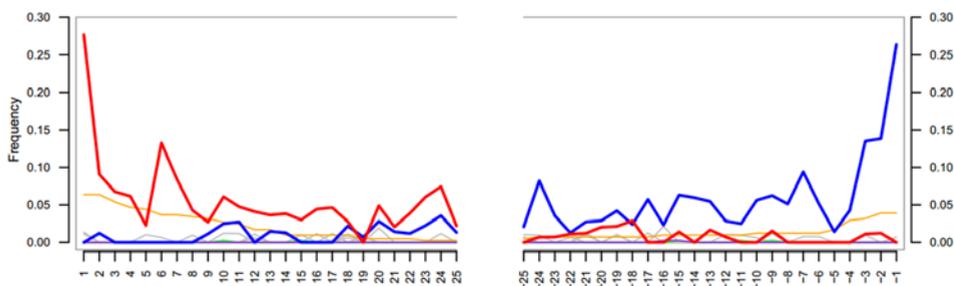
Successivamente, le sequenze che hanno effettuato il merging e che hanno passato i filtri di qualità vengono mappate sulla sequenza consenso tramite il software Burrows –Wheeler Aligner (BWA) utilizzando l'algoritmo MEM specifico per le sequenze più corte derivanti da sequenziamento con piattaforma Illumina. Il file risultante, avrà estensione .SAM. Quest'ultimo è un formato di testo che permette la memorizzazione di dati di sequenza in una serie di tabulazione delimitata in colonne in codice ASCII (Codice Standard Americano per lo scambio di informazioni). Questo file può essere inserito nel programma MapDamage 2.0 (Ginolach et al. 2011). MapDamage è uno strumento bioinformatico basato su una serie di scripts in Perl in grado di analizzare la distribuzione delle lunghezze dei frammenti e soprattutto di analizzare il pattern di frammentazione presente al 5'(CtoT) e al 3'(CtoA). Questo tipo di lesioni, sono tipicamente presenti alle estremità dei frammenti dell'aDNA, causate dalla deaminazione della Citosina. E' stato riscontrato sperimentalmente che nelle sequenze di DNA moderno questo pattern non è visibile oppure è presente in una percentuale inferiore al 5%, per cui MapDamage è uno strumento importantissimo per discriminare tra le sequenze endogene e quelle esogene (Briggs et al. 2007).

I risultati ottenuti con mapDamage della prima libreria (Fig.23) mostrano un pattern di degradazione piuttosto basso, di poco superiore al 5%. Questo indica la presenza di un debole segnale di antichità e della presenza di molti contaminanti.



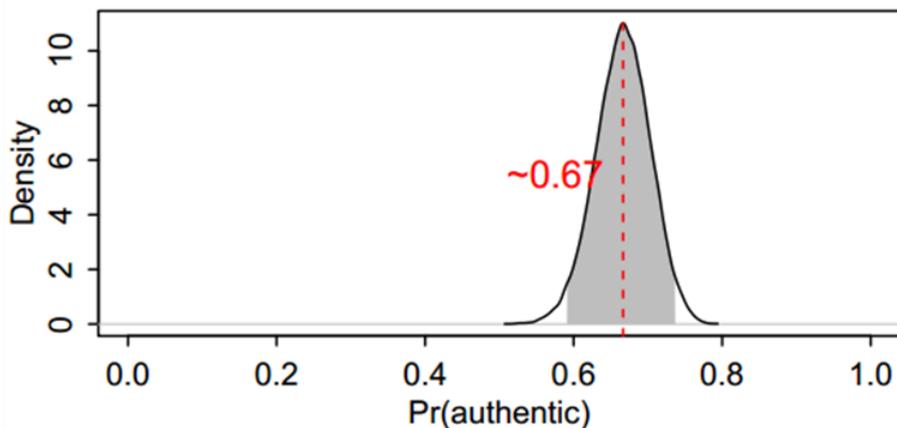
**Figura23:** Risultati di mapDamage inerenti alla prima libreria di Paus

Riguardo invece alla seconda libreria i risultati, mostrati in Fig.24, sono decisamente migliori. La percentuale del cambio C to T al 3' superano il 25% indicando una forte presenza del segnale di antichità. I risultati delle due librerie nel loro insieme ci indicano una situazione complessiva nella quale risulta evidente la presenza di sequenze endogene sovrastate però dall'elevata compresenza di contaminanti moderni.



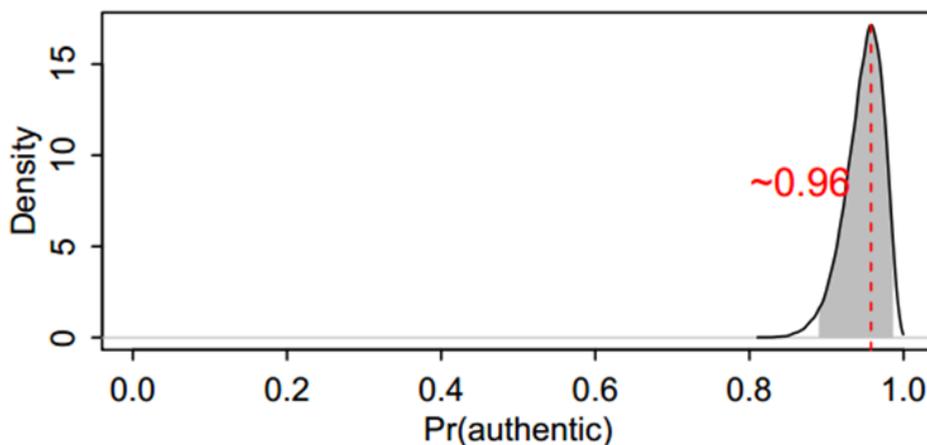
**Figura 24:** Risultati di mapDamage inerenti alla seconda libreria di Paus

Per dare ancor più robustezza ai risultati sono state condotte ulteriori analisi con il software contamMix (Fu et al.2013). Questo strumento bioinformatico tramite un approccio statistico bayesiano permette di stiminare la probabilità che le sequenze appartengono a dei contaminanti con un intervallo di confidenza del 95%. Più il risultato si avvicina ad 1 e più è alta la probabilità che le sequenze siano endogene. Il risultato di contamMix per la prima libreria di Paus (Fig.25) raggiungendo il valore di circa 0,67 si discosta notevolmente da uno indicando con notevole probabilità l'appartenenza delle sequenze a contaminanti. Diverso è il risultato di contamMix per la seconda libreria che invece con un valore di 0,96 si avvicina molto ad uno indicando il carattere endogeno delle sequenze (Fig.26).



**Figura 25:** Risultati di ContamMix inerenti alla prima libreria di Paus.

I risultati di contamMix nel loro insieme supportano i risultati con mapDamage per entrambe le librerie. Per la prima abbiamo infatti bassa percentuale di cambio C to T e un valore notevolmente distante da 1 per contamMix rafforzando quindi la supposizione che nella prima libreria di Paus la quantità di contaminanti sia predominante. Mentre invece per la seconda libreria i risultati dei due software coincidono indicando la presenza di sequenze provenienti principalmente da un'unica fonte biologica.



**Figura 26:** Risultati di contamMix inerenti alla seconda libreria di Paus.

Globalmente i risultati ottenuti dalle due librerie a doppio filamento indicano che nel reperto di Paus è ancora presente DNA endogeno. Il segnale di antichità infatti è chiaramente presente nella seconda libreria (misincorporazioni C/T >25%, Figura 24) mentre il basso numero di molecole effettivamente mappanti sulla sequenza di riferimento del genoma mitocondriale umano (Tabella 6 e 7) e i risultati contrastanti di ContaMix, indicano che nelle librerie è presente un'elevata quantità di DNA contaminante moderno che non permette di effettuare un'analisi delle sequenze tradizionale. Infatti le sequenze consenso ottenute dalle due librerie, che risultano parziali e a bassissima copertura, sono il prodotto di una miscela di sequenze autentiche e di sequenze contaminanti moderne e non sono pertanto informative ai fine della corretta caratterizzazione molecolare del reperto. Si è resa pertanto necessario una ripetizione della analisi con un approccio mirato a indagare meglio e caratterizzare il segnale di antichità presente nel campione.

### 5.3 Estrazione e libreria genomica a singolo filamento

Partendo da 20 ul di DNA estratto sono state fatte due librerie genomiche a singolo filamento, ognuna con 10 ul di DNA e senza trattamento UDG in modo da preservare alle estremità delle molecole il pattern di frammentazione, essenziale per la discriminazione tra sequenze endogene ed esogene. Il numero delle molecole per ogni libreria è stato determinato con la Digital Droplet PCR (Bio-Rad QX200), facendo una diluizione 1:5000 in buffer EBT del campione ed usandone 1 ul. L'indicizzazione con doppio indice delle librerie è stata effettuata usando AccuPrime Pfx DNA polymerase (Life Technologies), in 100 ul di reazione amplificando fino al raggiungimento del plateau di PCR. Successivamente 50 µL del prodotto di amplificazione sono stati purificati secondo il protocollo del MinElute PCR Purification Kit (Qiagen) ed eluiti in 20 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). I valori delle quantificazioni mostrate in Tabella 6 sono state ottenute con il Nanodrop.

	Polvere d'osso (mg)	Volume di DNA usato (µL)	Concentrazione droplet PCR	Numero di molecole totali	Concentrazione post-Indexing (ng/µL)
I° Libreria	25	10	2390	<b>1,20E+10</b>	325,94
II° Libreria	25	10	4370	<b>2,19E+10</b>	327,9

**Tabella 7:** La tabella mostra i valori della concentrazione del DNA nelle librerie genomiche e dopo l'indicizzazione

Dai risultati (Tabella 7) risulta che il protocollo messo a punto per le librerie a singolo filamento si è mostrato molto efficiente producendo buoni risultati

in termini di numero totali di molecole e concentrazione dopo l'indicizzazione. Come già precedentemente espresso, la procedura di indicizzazione si basa sul presupposto che superando il plateau di PCR la probabilità che si possano attaccare un elevato numero di indici aumenta.

### 5.3.1 Shotgun e Cattura del DNA mitocondriale

Inizialmente su un'aliquota delle librerie genomiche è stato eseguito un sequenziamento di tipo Shotgun esplorativo a bassa profondità. Fra tutte le reads prodotte sono state selezionate solo quelle di lunghezza maggiore di 35bp che mappano sul genoma umano di riferimento (hg19) (tabella 8). La percentuale di reads mappanti sul genoma umano è risultato piuttosto basso in entrambe le librerie, corrispondente a 1120 e 801 reads uniche rispettivamente.

	Raw	merged	filter_passed	L35	mapped	unique	% mapped
I Libreria	6.66E+05	6.13E+05	6.13E+05	4.40E+05	1.12E+03	1.12E+03	<b>0.26</b>
II Libreria	5.55E+05	5.25E+05	5.25E+05	3.50E+05	8.01E+02	8.01E+02	<b>0.23</b>

**Tabella 8:** Sequenze mappanti sul genoma umano

Successivamente è stata effettuata la cattura del genoma mitocondriale seguita da un sequenziamento a profondità elevata. In questo caso sono state ottenute 18,542,471 e 13,635,505 sequenze totali di cui 48,524 and 49,838 uniche, di queste circa il 65% e il 63% hanno mappato sulla sequenza di riferimento del genoma mitocondriale umano (tabella 9). Questi dati sottolineano come il DNA mitocondriale si preservi più facilmente anche in condizioni diagenetiche sfavorevoli o comunque complesse.

	Raw	merged	filter_passed	L35	mapped	unique	% mapped
I° Libreria	1.85E+07	1.76E+07	1.76E+07	1.38E+07	8.97E+06	4.85E+04	<b>65.17</b>
II° Libreria	1.36E+07	1.31E+07	1.31E+07	1.03E+07	6.51E+06	4.98E+04	<b>63.33</b>

**Tabella 9:** La tabella mostra la quantità di molecole dopo la cattura del DNA mitocondriale

### 5.3.2 Analisi Bioinformatica

La particolarità dell'analisi bioinformatiche effettuate con le sequenze ottenute sia dallo shotgun che dalla cattura del mtDNA nelle librerie a singolo filamento prevede l'applicazione di un filtro specifico per selezionare solo le sequenze che presentano misincorporazioni da Citosina a Timina alle estremità delle molecole.

Per quanto riguarda i risultati dello shotgun, l'analisi dei pattern di misincorporazione alle estremità della molecole ha restituito valori tra 0 e 4.3% (Tab 10) che si traducono in un numero di sequenze deaminate, e quindi effettivamente autentiche, molto basso pari a 13 e 9 reads per le due librerie, nessuna delle quali presenta cross-contamination a entrambe le estremità. Questi risultati nel complesso indicano una preservazione al livello di DNA nucleare estremamente bassa nel campione.

	1	2 C → T all sequences		3 C → T filtered sequences (cross-substitutions)	
	Filtered for C → T on 5' and/or 3' ends	5' end	3' end	5' end	3' end
	# sequences	frequency	frequency	frequency	frequency
I° Libreria	13	2.2 (6/272)	0 (0/132)	0 (0/0)	0 (0/2)
II° Libreria	9	0 (0/161)	4.3 (6/138)	0 (0/0)	0 (0/0)

**Tabella 10:** Risultati bioinformatici shotgun

Per quanto riguarda i risultati della cattura mitocondriale, nonostante il numero più elevato di sequenze anche in questo caso il pattern di misincorporazione alle estremità è risultato piuttosto basso, tra 2.1 e 3.2% (tabella 11). Tuttavia in questo caso, filtrando le sequenze in base alla presenza di misincorporazioni (1804 sequenze in tutto) e considernado solo le sequenze con cross-substitution, osserviamo un aumento nella frequenza di sostituzioni per arrivare a valori compresi tra 10.6 e 17.9% ovviamente con una diminuzione nei valori assoluti delle sequenze analizzate.

Questo dato conferma quanto evidenziato dal primo esperimento, ovvero la presenza di un segnale di antichità all'interno di una miscela di molecole endogene e di una grande quantità di molecole contaminanti.

	<b>1</b>	<b>2 C →T all sequences</b>		<b>3 C →T filtered sequences (cross-substitutions)</b>	
	<b>Filtered for C→T on 5' and/or 3' ends</b>	<b>5' end</b>	<b>3' end</b>	<b>5' end</b>	<b>3' end</b>
	<b># sequences</b>	<b>frequency</b>	<b>frequency</b>	<b>frequency</b>	<b>frequency</b>
<b>I° Libreria</b>	<b>906</b>	<b>2.3 (252/10946)</b>	<b>3.2 (207/6446)</b>	<b>10.6 (7/66)</b>	<b>16.7 (7/42)</b>
<b>II° Libreria</b>	<b>898</b>	<b>2.1 (218/10379)</b>	<b>3 (262/8728)</b>	<b>15.6 (10/64)</b>	<b>17.9 (10/56)</b>

**Tabella 11:** Risultati bioinformatici cattura del DNA mitocondriale

Il passaggio successivo delle analisi bioinformatiche è stato quello di andare a fare un'analisi differenziale delle sequenze in base alla presenza di posizioni diagnostiche neandertaliane. Dal confronto tra 11 genomi mitocondriali completi sequenziati e 311 sequenze mitocondriali moderne rappresentative della variabilità delle diverse linee mitocondriali attuali, è stato possibile determinare un elenco di posizioni in cui tutte le 10 sequenze neandertaliane differiscono dalle sequenze moderne. Tali posizioni sono 63 possono essere utilizzate come posizioni diagnostiche identificative di DNA mitocondriale neandertaliano (Fig.27).

H.Sapiens		H. Neanderthalensis	
243	A	G	246
417	G	A	426
438	C	T	447
2523	C	T	2560
2831	G	C	2868
3399	A	G	3439
3414	C	T	3454
3909	C	T	3949
4204	T	C	4244
4532	G	A	4572
4856	T	C	4896
4940	C	T	4980
5397	C	T	5427
5505	A	G	5545
5673	T	C	5713
5840	C	T	5883
6156	C	T	6206
6200	C	T	6250
6266	A	C	6316
6410	C	T	6460
6452	C	T	6502
6641	T	C	6691
7106	A	C	7156
7127	C	T	7177
7746	A	G	7797
7866	C	T	7919
7891	C	T	7942
8021	A	G	8072
8065	G	A	8116
8365	A	G	8417
8386	C	T	8438
8461	C	T	8513
8956	A	G	9008
9027	C	T	9089
9329	G	C	9401
9345	C	T	9417
9569	C	T	9641
10101	T	C	10173
10281	C	T	10353
10307	C	T	10379
10324	T	C	10396
10532	A	G	10604
10750	A	G	10822
11353	T	C	11455
11455	A	T	11550
11527	C	T	11599
11623	C	T	11695
11770	T	C	11842
11950	A	G	12022
12351	T	C	12423
12474	C	T	12547
13095	T	C	13168
13707	G	A	13780
13801	A	G	13874
13979	T	A	13952
15226	A	G	15299
15232	A	G	15305
15295	C	T	15368
15649	A	G	15722
15667	C	T	15740
16037	A	G	16113
16139	A	T	16218
16262	C	T	16346

**Figura 27:** Posizioni diagnostiche neanderthaliane (a sinistra) e la relativa posizione sulla rCRS

Andando a cercare queste posizioni nella totalità delle reads mappanti sul genoma mitocondriale ottenute da Paus si osserva che oltre il 98%

presentano la base moderna mentre solo poco più del 1% presenta la base neandertaliana (Tab. 12). Se la stessa analisi viene compiuta solo sulle sequenze filtrate in base alla presenza di deaminazione alle estremità, la percentuale di sequenze con la base neandertaliana sale globalmente al 20% (Tab. 12). Nelle sequenze deaminate, e quindi putativamente autentiche, il numero di sequenze con motivi esclusivi neandertaliani pertanto si arricchisce indicando che nel reperto di Paus è presente DNA antico autentico con caratteristiche neandertaliane.

	ALL SEQUENCES				FILTERED SEQUENCES			
	All humans (311) ≠ Neandertals (10)				All humans (311) ≠ Neandertals (10)			
	%Human	%Neandertal	observations	#observations	%Human	%Neandertal	#observations human	observations Neandertal
<b>PAUS I_LIB</b>	98,75	1,25	<b>7136</b>	<b>90</b>	73,86	26,14	<b>65</b>	<b>23</b>
<b>PAUS II_LIB</b>	98,63	1,37	<b>7492</b>	<b>104</b>	86,05	13,95	<b>74</b>	<b>12</b>
<b>Together</b>	98,69	1,31	<b>14628</b>	<b>194</b>	79,89	20,11	<b>139</b>	<b>35</b>

**Tabella 12:** Risultati dell' analisi bioinformatica considerando le mutazione private neanderthaliane.

Tale evidenza è rafforzata dal fatto che la percentuale di sequenze con base neandertaliana è maggiore e massima per le sequenze di dimensioni minori, comprese tra 35 e 45bp, dove raggiunge il 61,54%, mentre decresce all' aumentare della dimensione delle molecole (tabella 13).

	ALL SEQUENCES				FILTERED SEQUENCES			
	All humans (311) ≠ Neandertals (10)				All humans (311) ≠ Neandertals (10)			
Size bin	%Human	%Neandertal	# observations human	#observations Neandertal	%Human	%Neandertal	#observations human	#observations Neandertal
35-45	95,05	4,95	844	44	38,46	61,54	5	8
46-55	96,81	3,19	1154	38	66,67	33,33	12	6
56-65	98,14	1,86	1263	24	85,00	15,00	17	3
66-75	99,17	0,83	1432	12	71,43	28,57	15	6
76-85	98,60	1,40	1694	24	76,00	24,00	19	6
86-95	99,27	0,73	1634	12	88,24	11,76	15	2
96 and longer	99,40	0,60	6607	40	93,33	6,67	56	4
<b>All together</b>	98,69	1,31	<b>14628</b>	<b>194</b>	79,89	20,11	<b>139</b>	<b>35</b>

**Tabella 13:** Sequenze con posizioni neandertaliane analizzate in base alla lunghezza dei frammenti di DNA

## 6. DISCUSSIONE

Le diverse analisi genetiche effettuate sull'osso frontale umano arcaico rinvenuto nel 2009 sulle rive del fiume Po nei pressi di Cremona, hanno subito messo in evidenza l'estrema complessità del reperto.

I dati hanno evidenziato che la preservazione del materiale genetico endogeno del campione, denominato Paus, è molto limitata e parallelamente la contaminazione da DNA moderno esogeno, sia batterico che umano, è molto elevata. Questo ha reso le procedure molecolari effettuate in questa tesi, e volte a caratterizzare da un punto di vista filogenetico il reperto, molto complesse. Le prime analisi effettuate con un protocollo basato su NGS ma seguendo un approccio standard non specifico per campioni estremamente degradati e contaminati, ha messo in luce la presenza di un segnale di antichità nelle sequenze ottenute in una delle librerie, tuttavia data l'elevata contaminazione umana moderna che sovrasta il materiale genetico endogeno, non è stato possibile fare un'analisi filogenetica adeguata sui dati ottenuti.

È stata invece risolutiva l'analisi effettuata con il secondo approccio svolta in collaborazione con il Department of Evolutionary Genetics del Max Planck Institute of Evolutionary Anthropology di Leipzig, Germania, in particolare con il gruppo di Advanced DNA Sequencing Techniques guidato da Matthias Meyer. Tale approccio è stato utilizzato con lo scopo di favorire il recupero e la conversione in libreria genomica anche delle molecole di DNA di dimensioni molto piccole (<40bp). Inoltre l'approccio bioinformatico di analisi delle sequenze, prevede il filtraggio delle sequenze in base al fatto che presentino o meno gli specifici pattern di misincorporazione del DNA antico (Timina al posto di Citosina) alle estremità delle molecole. Con questo approccio combinato, oltre a massimizzare il recupero delle molecole endogene del campione (presumibilmente molto corte nei casi di scarsa preservazione), è possibile fare un'analisi differenziale delle sequenze ottenute separando le sequenze degradate, e quindi con maggiore probabilità endogene, dalle sequenze che non presentano degradazione. Le sequenze così separate sono state quindi indagate alla ricerca delle mutazioni diagnostiche del genoma mitocondriale neandertaliano rispetto ai mitocondri moderni. Con questa analisi, nonostante le poche molecole analizzate, è stato possibile mettere in luce che la percentuale di mutazioni diagnostiche neandertaliane è maggiore nelle sequenze degradate rispetto al pool unico di tutte le sequenze (20% contro 1% in media), e che questo trend è maggiore nelle molecole con dimensioni molto basse (tra 35 e 45bp).

I dati molecolari ottenuti hanno permesso quindi di dimostrare che nonostante le pessime condizioni di preservazione, l'osso di Paus contiene ancora tracce di DNA endogeno e che questo DNA presenta caratteristiche neandertaliane. Con i dati ottenuti e con il materiale biologico avanzato, non possiamo prevedere di fare al momento ulteriori analisi sia molecolari che

filogenetiche su questo reperto, perché non ci sono i presupposti per ricostruire il genoma mitocondriale completo o parziale ad un buon livello di coverage. Le evidenze delle analisi genetico-molecolari sono in accordo con i risultati dello studio morfologico tramite morfometria geometrica condotto dagli antropologi fisici dell'Università di Roma La Sapienza. Anche in questo caso l'analisi sul reperto è stata più complessa di quanto inizialmente pensato e ha richiesto la creazione di una database di confronto più ampio soprattutto in termini di campioni neandertaliani arcaici, ma i risultati finali delle analisi di morfometria geometrica hanno messo in evidenza che la morfologia di Pàus è compatibile con il range di variabilità dei neandertaliani.

Possiamo perciò concludere che il frontale umano di Paus appartiene a un Neandertaliano e questo risultato rappresenta la prima prova diretta che questi ominidi frequentassero nelle epoche glaciali anche le aree della pianura padana. Analisi future saranno volte a tentare di effettuare una datazione radiometrica del reperto in modo da contestualizzare in maniera sempre più precisa e approfondita questo eccezionale ritrovamento.

## 7 BIBLIOGRAFIA

- Adessi C, Matton G, Ayala G, Turcatti G, Mermod JJ, Mayer P, Kawashima E** 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res.* 28(20), E78
- Ávila-Arcos MC, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, Rasmussen M, Fordyce SL, 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 AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Pääbo S** 2009. Primer Extension Capture: Targeted Sequence Retrieval from Heavily Degraded DNA Sources. *J. Vis. Exp.* 31, 10
- Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S** 2007. Patterns of damage in genomic DNA sequences from a Neanderthal. *PNAS, USA.* 37,14616–14621
- Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S** 2007. Patterns of damage in genomic DNA sequences from a Neanderthal. *PNAS, USA.* 37,14616–14621
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Beaumont M, Barnett Ross, Austin J, Cooper A** 2007. Novel high-resolution reveals C>U-type base modification events as the sole cause of post-mortem miscodins lesions. *Nucleic Acids Res.* 35(17), 5717-28
- Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson, PLF, Xuan ZY, Rooks M, Bhattacharjee A, Brizuela L, Albert FW, de la Rasilla M, Fortea J, Rosas A, Lachmann M, Hannon GJ, Pääbo S** 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science.* 328, 723-725
- Caramelli D, Lari M** 2004. *Il DNA antico. Metodi di analisi e applicazioni.* Angelo Pontecorboli editore, Firenze
- Cooper A, Poinar HN** 2000. Ancient DNA: do it right or not at all. *Science.* 5482, 1139
- Dabney J, Meyer M, Pääbo S** 2013. Ancient DNA damage. *Cold Spring Harb Perspect Biol.* 1,5(7)
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M** 2013. Complete mitochondrial genome sequence of a Middle Pleistocene

- cave bear reconstructed from ultrashort DNA fragments. *PNAS, USA*. 110(39),15758-63
- Fortea J**, de la Rasilla M, Garcia-Taberner A, Gigli E, Rosas A, Lalueza-Fox C 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidron Cave (Asturias, Spain). *J Hum Evol*. 55, 353–357
- Gansauge MT**, Meyer M 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protoc*. 8(4),737
- García-Garcera M**, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, Lalueza-Fox C 2011. Fragmentation of contaminant and endogenous DNA in ancient sample determined by shotgun sequencing; prospect for Human paleogenomics. *PLoS ONE*. 6(8), e24161
- Gilbert MP**, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PGP, Germonprè M, Götherström A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrickson KM, Harkins TT, Sheridan S, Miller W, Shuster SC 2007. Whole Genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*. 28,1972-30
- Gilbert MTP**, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A 2003a. Characterization of genetic miscoding lesions caused by post-mortem damage. *Am J Hum Genet*. 72(1), 48-61
- Gilbert MTP**, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A 2003b. Distribution patterns of post mortem damage in human mitochondrial DNA. *Am J Hum Genet*. 72(1), 32-47
- Gilbert MTP**, Cuccui J, White W, Lynnerup N, Titball RW, Cooper A, Prentice MB 2004. Absence of *Yersinia pestis*-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology*. 150(2), 341-354
- Gilbert MT**, Wilson AS, Bunce M, Hansen AJ, Willerslev E, Shapiro B, Higham TF, Richards MP,
- O'Connell TC**, Tobin DJ, Janaway RC and Cooper A 2004a. Ancient mitochondrial DNA from hair. *Current Biology*. 14(12), R463-464
- Gilbert MTP**, Bandelt H-J, Hofreither M, Barnes I 2005a. Assessing ancient DNA studies. *Trends in ecology & evolution*. 20(10), 541-4
- Gilbert W**, Maxam A 1973. The Nucleotide Sequence of the Lac Operator. *PNAS, USA*. 70,3581-3584
- Ginolhac A**, Rasmussen M, Gilbert M. TP., Willerslev E, Orlando L 2011. MapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*. 27(15), 2153-5
- Green RE**, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K,

- Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S 2010. A Draft Sequence of the Neanderthal Genome. *Science*. 328(5979), 710-22
- Green** RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, Lachmann M, Pääbo S 2009. The Neanderthal genome and ancient DNA authenticity. *EMBO J*. 28, 2494–2502
- Green** RE, Malaspinas AS, Krause J, Briggs AW, Johnson PLF, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, Prüfer K, Siebauer M, Burbano HA, Ronan M, Rothberg JM, Egholm M, Rudan P, Brajković D, Kučan Ž, Gušić I, Wikström M, Laakkonen L, Kelso J, Slatkin M, Pääbo S 2008. A complete neanderthal mitochondrial genome sequence determined by high-throughput Sequencing. *Cell*. 134, 416-426
- Green** RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Pääbo S 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 444, 330-336
- Greenblatt** CL, Spigelman M 2003. *Emerging Pathogens the Archaeology, Ecology and Evolution of Infectious Disease*. Oxford University Press. Oxford
- Golenberg** EM, Giannasi DM, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature*. 344(6267), 656-8.
- Hansen** AJ, Willerslev E, Wiuf C, Mourier T, Arctander P 2001. Statistical evidence for miscoding lesions in ancient DNA templates. *Mol Biol Evol*. 18, 262-265
- Haack** W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Villems R, Renfrew C, Gronenborn D, Werner Alt K, Burger J 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*. 310, 1016-1018
- Handt** O, Höss M, Krings M, Pääbo S 1994. Ancient DNA: methodological challenges. *Experientia*. 50, 524-29
- Hänni** C, Brosseau T, Laudet V, Stheli D 1995. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleicacids research*. 23(5), 881-2
- Harris** DR 1996. *The Origin and spread of Agriculture and Pastoralism in Eurasia*. University College of London Press, London
- Hagelberg** E, Bell LS, Allen T, Boyde A, Jones SJ, Clegg JB 1991. Analysis of ancient bone DNA: techniques and applications. *Philos. Trans. R Soc Lond. B Biol Sci*. 333 (1268), 339-407

- Hofreiter M**, Jaenicke V, Serre D, Haeseler Av A, Paabo S 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids*. 29, 4793–4799.
- Hoss M**, Pääbo S, Vereshchagin NK 1994. Mammoth DNA sequence. *Nature*. 370,333
- Krause J**, Fu QM, Good JM, Viola B, Shunkov MV, Derevianko AP, Pääbo S 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*. 464, 894-897
- Krause J**, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, Pääbo S 2010. A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Current Biology*. 20, 231–236
- Lindhal T** 1993. Recovery of antediluvian DNA. *Nature*. 365, 700
- Lindahl T** 1993. Instability and decay of the primary structure of DNA. *Nature*. 362, 709-715
- Margulies M**, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Berka J, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E., Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM 2005. Genome sequencing in microfabricated highdensity picolitre reactors. *Nature*. 437, 376-380
- Maricic T**, Whitten M, Paabo S 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 5, e14004
- Meyer M**, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, Slatkin M, Reich D, Kelso J, Pääbo S 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 338(6104),222-6
- Meyer M**, kircher M 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc*. 10, 1101
- Mitchell D**, Willerslev E, Hansen A 2005. Damage repair of ancient DNA. *Mutat Res*. 571 (1-2), 265-76
- Mullis KB**, Faloona FA 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. 155, 335–350.

- Noonan** JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Dettler JC, Pääbo S, Rubin EM 2005. Genomic sequencing of Pleistocene cave bears. *Science*. 309, 597-600.
- Pääbo** S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M 2004. Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645-79
- Pääbo** S, Higuchi RG, Wilson AC 1989. Ancient DNA and the polymerase chain reaction. *The emerging field of molecular archaeology. The Journal of biological chemistry*. 264(17), 9709-12
- Pääbo** S 1989. Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification. *PNAS, USA*. 86, 1939-43
- Pääbo** S 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature*. 314, 644 - 645
- Poinar** HN, Carsten S, Qi J, Shapiro, MacPhee RDE, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC 2006. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science*. 5759,392-94
- Poinar** HN 2002. The genetic secrets some fossils hold. *Accounts of chemical research*. 35(8), 676-84
- Pruvost** M, Schwarz R, Correia VB, Champlot S, Braguier S, Morel N, Fernandez-Jalvo Y, Grange T, Geigl EM 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. *PNAS, USA*. 104(3), 739-44
- Reich** D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs U, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 468, 1053–1060
- Rollo** F, Marota I 1999. How microbial ancient DNA, found in association with human remains, can be interpreted. *Philos. Trans. R. Soc. Lond B Biol. Sci*. 334(1379), 111-9
- Rollo** F, Marota I 1995. Molecular phylogeny of the fungi of the Iceman's grass clothing. *Curr Genet*. 28(3), 289-97
- Rollo** F, Ashi W, Marota I 1994. Neolithic DNA from the Iceman's site: rare, long, and naked. *Ancient DNA new letter*. 3, 21-23
- Rizzi** E, Lari M, Gigli E, De Bellis G, Caramelli D 2012. Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution* 2012. 44, 21
- Saiki** RK, Gefland DH, Stoffe S, Scharf SJ, Higuchi R, Horn G, Mullis K, Erlich H 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239, 487-91

- Sampietro** ML, Gilbert MT, Lao O, Caramelli D, Lari M, Bertranpetit J, Lalueza-Fox C 2006. Tracking down human contamination in ancient human teeth. *Mol Biol Evol.* 23(9), 1801-7
- Sanger** F, Nicklen S, Coulson AR 1977. DNA sequencing with chain-terminating inhibitors. *PNAS, USA.* 74(12), 5463-7
- 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(3),e34131
- Sholtz** M, Giddings I, Push CM 1998. A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. *Anal Biochem.* 259(2), 283-6
- Stiller** M, Green RE, Ronan M, Simons JF, Du L, He W, Egholm M, Rothberg JM, Keates SG, Ovodov ND, Antipina EE, Baryshnikov GF, Kuzmin YV, Vasilevski AA, Wuenschell GE, Termini J, Hofreiter M, Jaenicke-Desprès V, Pääbo S 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *PNAS, USA.* 12, 103(37)
- Vigliardi** A 2002. La Grotta del Fontino. Una cavità eneolitica del grossetano. EDIFIR Edizioni Firenze
- Walden** KK, Robertson HM 1997. Ancient DNA from amber fossil bees?. *Mol Biol Evol.* 14(10), 1075-7
- Willerslev** E, Cooper A 2005. Ancient DNA. *Proc. R Soc. B.* 272
- Willerslev** E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, Cooper A 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science. Epub.* 300(5620),791-5
- Woodward** SR, Weyand NJ, Bunnell M. DNA sequence from Cretaceous period bone fragments. *Science.* 5188, 1229-1232
- Yang Dongya** Y, Watt K 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. *Journal of Archaeological Science.* 32, 331-336
- Zvelebil** M 2004. Who were we 6,000 years ago? In search of prehistoric identities. *Traces of ancestry: studies in honour of Colin Renfrew.* Cambridge, UK: Mc Donald Institute Mnographs. 41-60
- Papi, F. &