



Antropologia

Biologia molecolare applicata (Università degli Studi di Firenze)



Scansiona per aprire su Studocu

ANTROPOLOGIA

Lezione 1 28.09

Antropologia affronta lo studio dell'uomo nella sua qualità di specie zoologica, in particolare la storia naturale dell'uomo attraverso le molecole (DNA). In particolare sviluppa i seguenti argomenti: L'antropologia come Storia Naturale dell'Uomo. Tappe dell'evoluzione umana. Le basi genetiche dell'evoluzione. Aspetti biomolecolari per lo studio dei processi evolutivi. Marcatori genetici. Azione dell'ambiente sull'evoluzione delle popolazioni e delle specie. Il concetto di specie e evoluzione della specie. La paleogenetica. Caratteristiche del DNA antico. Principali tecniche biomolecolari nello studio di reperti antichi. Metodologia classica e il Next generation Sequencing. I marcatori molecolari negli studi sul DNA antico. Come si effettua uno studio sul DNA antico. Le Forme fossili.

Gli studi dell'antropologia si possono dividere in **Pre-DNA era** e **Post-DNA era**

Pre-DNA era: ovvero prima del 1953, anno della scoperta della struttura del Dna, si studiavano principalmente le proteine degli individui viventi. Come? Si prelevava il sangue da primati (scimpanzè, gorilla, uomo) e vedevamo quali erano affini tra loro, confrontando le differenti cariche degli amminoacidi che costituivano le proteine degli organismi studiati. Si estraevano le proteine, venivano messe in un campo elettrico, venivano fatte migrare e a seconda della migrazione delle proteine/del loro punto isoelettrico/della carica amminoacidica si vedeva che proteine simili andavano dalla stessa parte. Avendo un sistema proteico più simile allora forse avranno avuto un antenato vicino in passato e quindi venivano costruiti i primi alberi genetici

Post-DNA era: Con lo sviluppo di nuove tecnologie legate soprattutto alla affermazione di nuove tecniche di biologia molecolare (PCR, Sequenziamento automatico, MicroArray, Sequenziamento NGS) si assiste ad un notevole incremento dei lavori che indagano direttamente le sequenze nucleotidiche del DNA, indagini che permettono di chiarire numerosi aspetti filogenetici, evolutivisti e popolazionistici.

Grazie allo sviluppo di queste nuove tecniche è dunque possibile estendere gli studi non solo sul vivente ma anche sui reperti fossili: vengono pubblicati nella seconda metà degli anni ottanta i primi studi sul DNA antico (aDNA). Reperti fossili di 100-500 a 20000 anni fa → ricostruendo la storia evolutiva di HOMO.

Nella costruzione di un genoma si possono trarre tante informazioni interessanti: variabilità genetica ma anche a livello fenotipico (capelli, colore occhi), ma anche a livello fisico (vertebre, scapole).

Denny: ibrida: genitori denisoviano (papa) e neandertaliano (mamma)

Il corso si baserà sull'analisi e studio del genoma umano, in particolare la struttura del genoma umano.

Il termine **genoma umano** designa l'insieme di tutte le molecole di DNA che sono presenti in ogni cellula del corpo umano eccettuato gli eritrociti. Non tutto il genoma è nel nucleo (**nDNA**), esiste una parte molto importante contenuta nei mitocondri e chiamata DNA mitocondriale (**mtDNA**)

Il DNA mitocondriale sarà il primo ad essere studiato: quindi dalla variabilità genetica individuale.

Il DNA contiene tutte le informazioni necessarie per lo sviluppo dell'organismo e queste informazioni sono contenute nei geni.

La cosa sorprendente è che i geni costituiscono solo il 3% del DNA totale (circa 20.000). Il resto non è codificante ma svolge un ruolo funzionale nella regolazione e nella promozione dell'espressione genica o un ruolo strutturale nell'integrità dei cromosomi durante la divisione nucleare.

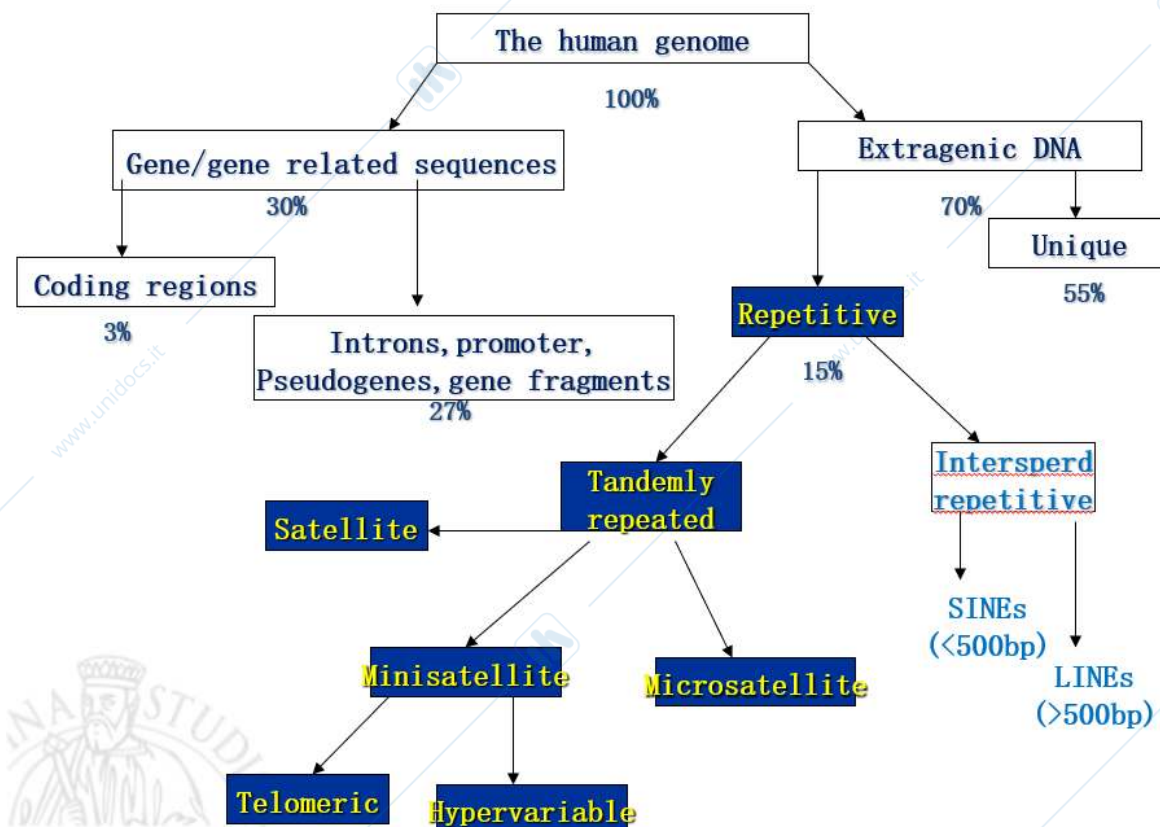
Ci focalizzeremo sul DNA non codificante perché non è sottoposto a fenomeni selettivi. Ma viene modificato dal tempo e caso.

Il progetto genoma ha permesso di conoscere più in dettaglio l'architettura delle sequenze di DNA

Il 15, ed il 16 febbraio 2001, le riviste *Nature* (<http://www.nature.com/nature/index.html>) e *Science* (<http://www.sciencemag.org/>) pubblicano l'intera sequenza del genoma umano.

La cosa sorprendente è che i geni risultano essere in numero decisamente inferiore a quanto stimato tra intorno ai 23.224. Appena il triplo della drosophila circa il doppio di quelli di un anellide *Caenorhabditis*.

STRUTTURA GENOMA



Sequential breakdown of the genome into component DNA types (Bennett P: Demystified...Microsatellites. J Clin Pathol Mol Pathol 2000,53:177-183)

Differenza tra microsatellite, minisatellite e satellite è nella grandezza delle ripetizioni.

Per i nostri studi il DNA che interessa nel DNA extragenico, un po nel codificante, e microsatellite.

DNA RIPETITIVO INTERSPERSO

Le singole unità ripetute non sono raggruppate ma sparse in più punti del genoma. Gli esempi più comuni sono le sequenze SINEs (Short Interspersed Nuclear Element) e LINEs (Long Interspersed Nuclear Element)

LINEs

- Sono associate prevalentemente a DNA genomico ricco in A/T perché tendono a posizionarsi in regioni del cromosoma povere di geni allo scopo di imporre il minimo impatto mutazionale al genoma.

SINEs

- Sono associate prevalentemente a DNA genomico ricco in G/C. *Perché?* Sembra che tali sequenze svolgano una qualche funzione positiva per il genoma: esse sarebbero espresse in condizioni di stress ed i risultanti RNA legherebbero una particolare protein chinasi PKR e bloccherebbero la sua capacità di inattivare la traduzione.

DNA RIPETUTO IN TANDEM

DNA costituito da blocchi di sequenze che possono trovarsi in poche o in molte localizzazioni cromosomiche differenti. A seconda delle dimensioni medie del blocco di unità ripetute è possibile suddividere questo DNA non codificante in subclassi: satellite, minisatellite e microsatellite

DNA SATELLITE: comprende un grande gruppo di blocchi di DNA ripetuto in tandem con singole unità ripetute più o meno complesse. Il DNA di questo tipo non è trascritto e praticamente costituisce il grosso dell'eterocromatina del genoma localizzata al centromero (eterocromatina pericentrica), intersperso nei cromosomi.

DNA MINISATELLITE comprende due tipi di sequenze di DNA ripetuto in tandem, non trascritto, di modeste dimensioni e disperso nel genoma nucleare: minisatelliti telomerici e minisatelliti ipervariabili.

DNA TELOMERICO

- Questo tipo di DNA è localizzato all'estremità dei cromosomi, nei telomeri
- Il costituente principale del DNA telomerico è rappresentato da 10-15kb di unità esanucleotidiche ripetute in tandem, in particolare TTAGGG, che vengono aggiunte da un enzima specializzato, la telomerasi
- Tali ripetizioni sono responsabili della funzione dei telomeri in quanto agiscono come protezioni delle estremità dei cromosomi dalla degradazione e dalle perdite di materiale

- Inoltre forniscono un meccanismo per la replicazione delle estremità lineari del DNA cromosomico

DNA MINISATELLITE IPERVARIABILE o VNTR (Variable Number of Tandem Repeat)

- Sono sequenze altamente polimorfiche ed organizzate in oltre 1000 gruppi (lunghi da 0.1 a 20 kb) di corte unità ripetute in tandem, che variano considerevolmente per dimensioni ma posseggono una sequenza comune centrale (*core*),GGGCAGGAXG
- Molti di questi gruppi si trovano vicino ai telomeri
- La maggior parte di queste sequenze *non* sono trascritte eccetto alcuni elementi all'interno di sequenze intrageniche non codificanti
- Il significato non è ancora chiaro, ma indipendentemente dalla loro reale funzione nel genoma umano, esiste un utilizzo pratico di questi gruppi, ma in genere delle sequenze ripetute, nel DNA fingerprinting (impronta digitale del DNA) → perché sono ipervariabili tra un individuo all'altro.

DNA FINGERPRINTING

La molecola del DNA possiede caratteristiche che consentono di differenziare i diversi individui perché, diversamente da un'impronta digitale convenzionale, l'impronta digitale del DNA non può essere alterata da alcun trattamento conosciuto. È vero, ma con qualche manovra di carattere biologico può ulteriormente degradarlo e difficile da leggere

Queste differenze vengono utilizzate per studi che trovano sviluppo in diversi ambiti:

Forense: risoluzione dei crimini

identificazione vittime

Medico: diagnosi dei disordini ereditati

cure di sviluppo per i disordini ereditati

test di paternità

Evoluzionistico: comportamento degli animali

popolazioni genetiche

DNA MINISATELLITE IPERVARIABILE: PROBLEMI

I VNTR presentano delle difficoltà relative alla genotipizzazione in quanto vista la loro lunghezza tali marcatori vengono amplificati con difficoltà in una reazione di PCR

Non sono uniformemente distribuiti in tutto il genoma

Quindi per Dna fingerprinting viene utilizzato il DNA microsatellite (piu piccolo e facilmente analizzabile)

LEZIONE 2 29.09

ALTRA CATEGORIA DI MARCATORI MOLECOLARI: **DNA MICROSATELLITE O STR** (short tandem repeat)

Molto utilizzati in ambito forense perché è POLIMORFICO.

Le famiglie di DNA microsatellite includono piccoli blocchi di ripetizioni in tandem, da 15-50 volte, semplici nella sequenza (1-6bp) e dispersi nel genoma.

Sono numerosi, con una densità media stimata intorno ad 1 STR ogni 30,000-10,000bp.

Il numero di ripetizioni varia frequentemente tra gli alleli: genera alto grado di **polimorfismo**

STR sono POLIMORFI: presenti in differenti forme in differenti individui: avere molti varianti in un gruppo di persone permette di utilizzarli in maniera piu dettagliata come DNA fingerprinting in quanto CARATTERIZZA UNA PERSONA.

Es. Classe da 10-15 persone posso trovare differenti variante in ciascuna persona

I VNTR erano utilizzati come DNA fingerprinting ma ci sono delle problematiche nell'analizzarli: sono piu lunghi, meno frequenti e quindi piu complessi da analizzare.

Sono abbastanza **comuni le ripetizioni mononucleotidiche (A)_n o (T)_n** che arrivano a costituire in tutto ca 10 Mb, ovvero lo 0.3% del genoma nucleare

Le ripetizioni di (G)_n e di(C)_n sono, al contrario, più rare

Le ripetizioni dinucleotidiche **CA/TG** (TG sul filamento complementare) sono molto comuni; costituiscono infatti ca lo 0.5% del genoma

Sono *altamente* polimorfiche

Sono diffuse anche le ripetizioni **CT/AG** che si verificano in media ogni 50 kb e costituiscono ca lo 0.2% del genoma

Le ripetizioni CG/GC risultano molto rare, questo perché i residui C affiancati in 3' da un residuo G (cioè CpG) sono soggetti a metilazione e a successiva deaminazione, diventando TpG (o CpA sull'altro filamento)

I trinucleotidi ed i tetranucleotidi ripetuti in tandem sono relativamente rari ma spesso polimorfici e potenzialmente utili per l'identificazione di marcatori altamente polimorfici

I microsatelliti sono evidenziabili non solo nel DNA intergenico o dentro gli introni ma esempi ne sono stati riportati anche all'interno di regioni codificanti (ma non nell'uomo) ad es. a livello mitocondriale.

VANTAGGI:

Tecnicamente gli STR, a confronto con altri marcatori, presentano il vantaggio di essere:

1. Numerosi
2. Equamente distribuiti in tutta la parte eucromatica del genoma
3. Altamente polimorfici e quindi informativi
4. Analizzabili facilmente tramite PCR
5. Questa relativa facilità di studio ha portato ad un rapido incremento del numero degli STRs esaminati

Come si identificano gli STR

Catalogazione standard: es D12S324, dove 12 è il cromosoma sul quale il marker è localizzato e 324 è un codice identificativo

(Bennett P: Demystified...Microsatellites. *J Clin Pathol: Mol Pathol* 2000,53:177-183)

Individuo eterozigote

5' ca ca ca ca ca ca ca ca ca ca ca ca ca ca ca 3'
5' ca ca ca ca ca ca ca ca ca ca ca ca ca ca 3'

Individuo omozigote

5' ca ca ca ca ca ca ca ca ca ca ca ca ca ca 3'
5' ca ca ca ca ca ca ca ca ca ca ca ca ca ca 3'

Aspetto polimorfico STR:

individuo omozigote: sullo stesso locus ha sequenze uguali

Individuo eterozigote: sullo stesso locus ha sequenze diverse.

Gli STR sono polimorfici, significa che a ciascun locus ci possono essere differenti forme di quel

determinato frammento di DNA, cioè determinati alleli.

Prendo due individui: un eterozigote ha Ca 17 su un allele e Ca 13 sull'altro. Un omozigote Ca 13 in entrambi gli alleli. Quindi questi differiscono per Ca 17.

Se prendo un altro individuo eterozigote a quel determinato locus è Ca 22 e Ca 15.

Altro individuo omozigote: Ca 12 e Ca 12. Quindi ho una certa variabilità in base al numero di alleli che vado a ricercare nella popolazione che vado ad analizzare che sono polimorfici.

Questo è molto importante perché con questo polimorfismo posso fare un riconoscimento individuale.

Suppongo che sono eterozigote Ca 17 Ca 13. Sicuramente ho ricevuto Ca 17 da mio padre o da mia madre e Ca 13 da mio padre o da mia madre, quindi per saperlo devo andare a ricercare tali alleli a quel determinato locus.

Se ho un fratello o una sorella: questi possono avere i polimorfismi che hanno ereditato insieme a me secondo le leggi mendeliane.

Il fatto che questi alleli siano polimorfici mi permette di andare a costruire un profilo individuale all'interno di un individuo andando ad analizzare più alleli in differenti loci.

Nel caso dovessi fare un riconoscimento di paternità posso confrontare STR presenti nell'individuo e STR presenti nel padre.

Quindi si può utilizzare per identificazioni personali.

In una scena del crimine.

Per ricostruire relazioni parentali in sepolture multiple negli etruschi (in ambito archeologico)

In ambito medicale: come marker tumorale → perché la quantità di STR con il numero di cellule tumorali aumentano

DOMANDA: quanti alleli possono essere presenti in ciascun locus degli STR? In un locus potremmo avere infiniti alleli (non si sa quanti perché si dovrebbe analizzare tutta la popolazione)

ALTRA CATEGORIA DI MARCATORI MOLECOLARI: utilizzati negli studi di caratteri evolutivo.

SNP (single nucleotide polymorphism): sia DNA genico sia DNA extragenico:

Ci sono 4 varianti. Transizioni più frequenti rispetto transfezioni.

- Gli SNPs, scoperti negli anni 80, sono variazioni di sequenza del DNA che si verificano quando è alterato un singolo nucleotide della sequenza genomica
- Affinchè una variazione nella sequenza nucleotidica sia considerata uno **SNP, deve essere presente in almeno l'1% della popolazione** per cui l'inserzione/delezione di una singola base non deve essere considerata uno SNP
- A differenza delle VNTR e degli STRs, gli SNPs non sono sequenze ripetute e possono trovarsi sia nelle regioni codificanti che non-codificanti del genoma (Dwight H.O. et al. *Journal of Molecular Diagnostics*, 2000, 2(4):202-208)

Mentre gli STR si trovano in parti non codificanti del genoma, dal punto di vista "selettivo" non hanno influenza., rispetto agli SNP perché possono trovarsi sia in regioni codificanti e non codificanti. Quando vengono studiati, mi devo porre il problema di quali SNP andare a cercare e dove andare a cercare (nella regione codificante e reg non codificante).

Es. studio di una popolazione; cioè capire la storia evolutiva di una popolazione mi devo porre il problema di quali SNP andare a cercare e dove andare a cercare (nella regione codificante e reg non codificante). Ciò dipende dal tipo di studio da fare.

Esempio: persone con pelle chiara. Un antropologo dice che la nostra origine non è un'origine con pelle chiara. Sicuramente siamo una specie che viene dall'Africa e intorno a 90000 anni fa quando siamo usciti dall'Africa sicuramente la nostra pelle era scura.

Pelle scura è un adattamento alla luce del sole perpendicolare.

Ora gli europei hanno la pelle chiara: la selezione ha avuto un ruolo importante:

Ci sono geni che codificano per il colore della pelle: MC1R che a seconda delle varianti di questo gene può essere espressa in modo diverso la quantità della melanina

In Africa avere la pelle scura è un vantaggio, in Europa è uno svantaggio, perché il sole non è diretto ma meno incidente e avere la pelle scura non permette di sintetizzare vitamina D.

Chi aveva la pelle molto scura ha avuto una fitness minore, chi aveva la pelle un tantino più chiara aveva una fitness maggiore →

Si traduce in una variabilità genetica in questi individui. In questo caso la selezione ha favorito gli alleli di individui con pelle più chiara → che hanno una fitness maggiore.

Se faccio quindi uno studio sulla variabilità genetica del gene MC1R, questo studio è viziato dalla selezione naturale che agisce su questi geni.

Per capire se due popolazioni derivano da una popolazione (es. africana quindi pelle scura) devo ricercare geni che non codificano per nessun tipo di proteina, quindi la parte extragenica.

Vado a cercare gli SNP che stanno nel DNA extragenico.

La maggiore parte di questi studi popolazionisti utilizzano questi SNP che stanno nelle regioni extrageniche

- Attualmente sono i marcatori molecolari più utilizzati perché il grande vantaggio nell'utilizzarli è dato dall'elevato numero di polimorfismi che possono essere genotipizzati e dalla loro elevata densità lungo tutto il genoma
- **Costituiscono ca il 90% di tutti i polimorfismi** presenti nel genoma umano. A giugno del 2004 nell'uomo è stata stimata una frequenza per gli SNP pari a 1/700bp
- **Il recente progresso della genomica ha messo in luce come una parte rilevante della variabilità tra individui sia da attribuirsi a polimorfismi a singolo nucleotide**

Senza variabilità genetica non ci sarebbe evoluzione.

- Gli SNPs acquistano particolare rilevanza in campo biomedico quando possono essere messi in relazione a patologie che non presentano una trasmissione genetica semplice
- Confrontando lo schema e le frequenze degli SNP su geni potenzialmente coinvolti in patologie e i fenotipi esibiti dai soggetti portatori, è possibile utilizzare tali sequenze come marcatori molecolari

PROBLEMI

- Attualmente sono poco conosciute le distribuzioni degli SNP all'interno di diverse popolazioni
- Non essendo uguali tutti gli SNP, per capire il loro effetto sarà importante eseguire un'analisi computazionale prima di eseguire uno studio relativo al loro eventuale coinvolgimento in una patologia

Popolazione africana conserva la maggior parte della variabilità genetica.

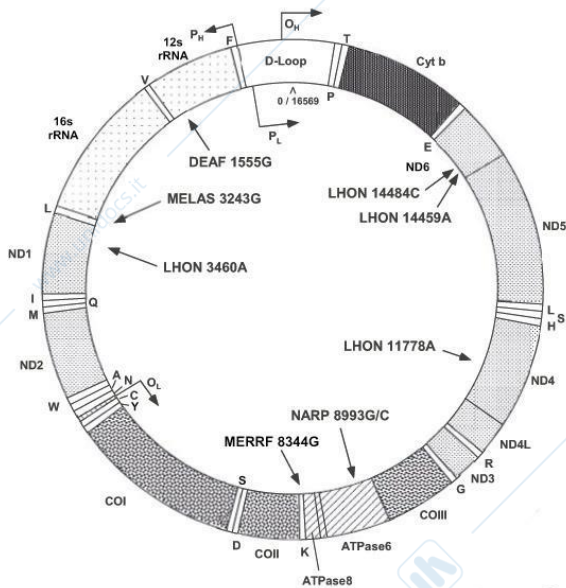
INDIVIDUAZIONE DEGLI SNP

- Sequenziamento diretto dei prodotti di PCR di regioni genomiche contenenti il gene di interesse in individui diversi
- Tecnologie NGS

Il DNA mitocondriale: altro MARCATORE MOLECOLARE

A differenza di quello nucleare è una molecola circolare si trova nei mitocondri (organelli cellulari) che svolgono un ruolo di primaria importanza nella respirazione cellulare.

Molecola dell'mtDNA 16569(O 59) bp :



Contiene geni che codificano per

- RNA ribosomiali,
 - geni che codificano per 22 RNA transfer indicati con le singole lettere dell'amminoacido corrispondente
 - geni che codificano per 13 proteine → 3 sub unità della citocromo ossidasi, 7 sub unità della NADH deidrogenasi e 2 sub unità della ATPasi ed il citocromo b.
- 2 CATENE: H (pesante) e L (leggera), dipende dal quantitativo di GC

Non possiede introni

C'è un regione D-loop non codificante: che significa ansa di dislocazione da dove parte la replicazione del filamento L → interessanti dal punto di vista evolutivo.

Caratteristiche non strutturali:

Il DNA mitocondriale (mtDNA) è uno dei marcatori genetici più utilizzati negli studi popolazionistici e filogenetici: quindi è molto usato negli studi di antropologia molecolare

Caratteristiche:

- Si eredita per sola via materna: si può utilizzare la molecola come strumento di riconoscimento individuale. Da un punto di vista evolutivo posso ricostruire la storia evolutiva di una specie.

- Non partecipa a fenomeni di ricombinazione: per questo si utilizza come marcatore genetico per ricostruire la storia evolutiva di una specie. Se il DNA mitocondriale ricombinava, differenze in un segmento potevano essere dovute ad un passaggio meccanismo (crossing over) e non ad un passaggio nel tempo.
- Ha un elevato tasso di mutazione stimato $4,7 \cdot 10^{-9}$ sost/ sito/ anno → si può utilizzare come orologio molecolare per sapere dopo quanto tempo si sono separati dalla progenitrice → un determinato mitocondrio ha una frequenza maggiore.

Es. una popolazione che vive in un equilibrio di Hardy-weinberg, cioè una popolazione non isolata, che scambiano tra loro i geni. Per vari fenomeni (accoppiamento, rimescolamento casuale), se la popolazione non è molto grande posso trovare varianti particolari di DNA mitocondriale: si affermerà una linea mitocondriale rispetto ad altre → deriva genica. Si osserva quindi un maggior quantità di un determinato allele.

In questa popolazione ad un certo momento degli individui decidono di migrare da un'altra parte: alla prima generazione avranno gli stessi mitocondri da cui si sono generati. Col passare delle generazioni a causa di un tasso di mutazione abbastanza elevato, si hanno mutazioni che si fissano nelle due popolazioni e quindi le differenze cominciano ad aumentare.

Se l'antropologo che fa un'analisi delle due popolazioni: analizzando il DNA mitocondriale si vede che sono abbastanza simili quindi suppone che potrebbero derivare da una stessa popolazione.

Quanto tempo fa è avvenuta questa separazione dalla popolazione madre? Conto le sostituzioni nucleotidiche. Se la sostituzione avviene ogni x anni, andando a contare le sostituzioni so da quanto tempo si sono separate. Perciò viene utilizzato come orologio molecolare.

Quindi tutte queste caratteristiche ci permettono di utilizzarlo negli studi di carattere evolutivo e popolazionistico

- Esiste in un elevato numero di copie per cellula: probabilità elevata di ritrovare un DNA mitocondriale all'interno di un individuo vissuto nel passato rispetto al DNA nucleare.

Infatti i primi studi fatti sul DNA degradato sono studi fatti ottenuti dai mitocondri.

CROMOSOMA Y: si eredita per via paterna, ha caratteristiche simili al DNA mitocondriale. Usato per studi di carattere filogenetico popolazionistico perché ha un'eredità esclusivamente paterna.

Quindi cromosoma Y e DNA mitocondriale → Due strumenti che ci permettono di descrivere la storia di una popolazione in due linee differenti: la linea al femminile e la linea al maschile.

Il DNA nucleare ci descrive l'intera storia di una popolazione o di una specie.

APLOTIPI E APLOGRUPPI

APLOTIPI

combinazione di diversi stati allelici di un set di marcatori polimorfici che si trovano fisicamente associati sulla stessa molecola di DNA, per esempio un cromosoma o una regione cromosomica

Aplogruppo:

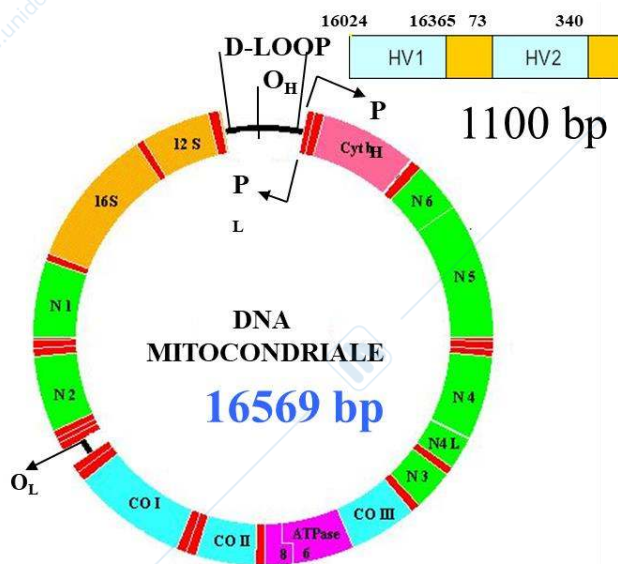
gruppo di aplotipi di cui si ipotizza un'origine comune, grazie alla condivisione di mutazioni caratteristiche, generalmente ad evoluzione lenta.

Cerchiamo di declinare le due caratteristiche sul DNA mitocondriale.

Quindi aplotipo è caratterizzato da mutazioni specifiche che stanno su una regione cromosoma. Il dna mitocondriale essendo una regione unica che può avere delle mutazioni particolari.

Indificazione aplogruppi mitocondriali: si vanno ad osservare mutazioni presenti in D-loop e mutazioni presenti nella regione codificante. Si sono accorti che sulla regione non codificante (D-loop) possono essere presenti delle mutazioni che caratterizzano questa regione

In D-loop ci sono delle regioni chiamate HV1 e HV2 che sono regioni ipervariabili. Vuol dire che possono presentare molte mutazioni. Sono mutazioni a singolo nucleotide (SNP).



Studiando il DNA mitocondriale si osservavano mutazioni nelle due regioni (codificanti e non codificanti). Individui che appartenevano alla stessa pop o stesso gruppo condividevano le stesse mutazioni nella reg codificante e reg non codificanti (nativi americani) specifiche e differivano da individui di popolazione diversa (europei).

Portò quindi al raggruppamento dei mitocondri che possedevano gli stessi motivi mutazionali e questi gruppi differivano tra individui di popolazione differente.

Un aplogruppo mitocondriale: si caratterizza da mutazioni su regioni non codificante e codificante del mitocondrio.

È un insieme di aplotipi che presentano le stesse mutazioni nella reg codificante e non codificante del mitocondrio.

Essendo il DNA variabile non tutte le mutazioni saranno condivise ma un buon numero di mutazioni saranno condivise tale che due individui avendo aplotipi differenti appartengono allo stesso aplogruppo.

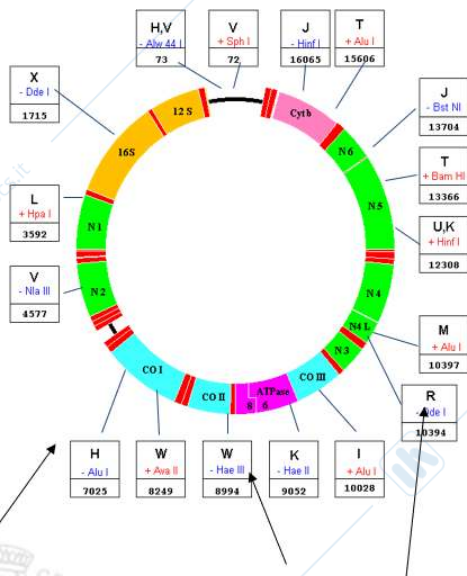
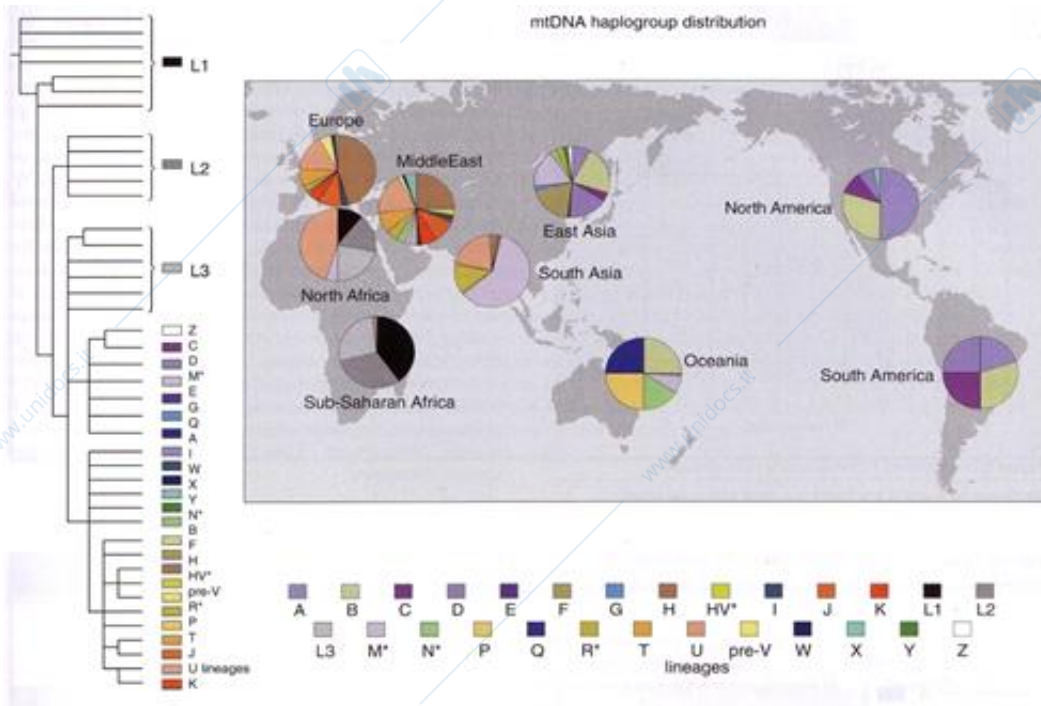
→ È possibile ricostruire i movimenti migratori: sono studi di filogeografico.

DOPPIA ELICA CIRCOLARE CONTENENTE 16569 COPPIE DI BASI

-due filamenti che lo costituiscono sono denominati **H (heavy)** e **L (light)** per la differente costituzione nucleotidica:

H contiene in prevalenza **basi puriniche A e G** (di peso molecolare maggiore)

L : basi pirimidiniche **C e T** (di peso molecolare minore).



Ogni aplogruppo comprende più **APLOTIPI** definiti sulla base della presenza di posizioni mutanti nella regione d-loop; tali posizioni sono chiamate **caratterizzanti** quando rappresentano quelle tipiche dell' aplogruppo.

Eteroplasmia= aplotipi multipli nello stesso individuo
Omoplasmia= un solo tipo di mtDNA nello stesso individuo

APLOGRUPPI- Vengono definiti in base alla **presenza o all' assenza di sostituzioni nucleotidiche** in determinate posizioni della porzione codificante del genoma mitocondriale

FILOGENIA: Struttura ad albero che rappresenta le relazioni evolutive tra insieme di Taxa

FILEOGEOGRAFIA:

Analisi della distribuzione geografica di diversi cladi all'interno di una filogenia. La filogenia fornisce una dimensione temporale ed evolutiva che è combinata con la dimensione spaziale della geografia

Gli aplogruppi mitocondriali possono essere utilizzati per ricostruire i movimenti migratori: studi di carattere filogeografico e filogenetico

150 mila anni fa la situazione che si osserva

Aplogruppi mitocondriali L0 in Africa

Prima espansione dall'africa verso tutto il continente

Da L0 si origina L1 → diffusione di popolazione sapiens in africa intorno a 60 80 mila anni fa:
L2 e L3 → espansione e ripopolamento del continente africano

60-65-70 mila anni fa: fuoriuscita di Homo sapiens dall'africa: sviluppo aplogruppo M e N nella porzione medioorientale

40 mila anni fa: Nuova espansione verso Australia, Asia meridionale, India e Asia centrale: M e N.

Dagli aplogruppi MN → a b F D M Asia

Da aplogruppi M e N in Europa → H U I J T, dove incontrano i neandertaliani

15-18 mila anni fa → Dal continente asiatico attraversano lo stretto di Bering e vanno in America

A B C D X

Anni 20 mila anni fa si assiste ad una glaciazione → Popolazioni che abitavano verso i poli si ritirano verso la zona centrale dove il clima è più temperato.

Nuova espansione. Risenti

Con questi movimenti si sviluppa aplogruppo V, Z in Asia, Y sempre in Asia

Popolazione che aveva lasciato l'Africa, successivamente sono ritornati in Africa (senti)

1500 2000 anni Cristo

Popolazione Isola del Pacifico con aplogruppo B



situazione attuale

senti

Inevitably, genetic research and archaeological discoveries will modify the conclusions presented here, and some details have been knowingly simplified for presentation. Major technical stages in this project involved the development of phylogenetic networks to reconstruct the mtDNA tree, the refinement of the mtDNA mutation rate as a molecular clock for dating prehistoric events, and last but not least, the systematic detection and elimination of published (and our own!) data errors. All these aspects remain part of our research activities.

UN PO DI STORIA

articolo 1984: DNA sequences from the quagga an extinct member of the horse family

Perché DNA degradato e non DNA antico? Perché la degradazione non dipende dall'antichità

Sono i primi due articoli che trattano di questa disciplina → studi di carattere antropologico

Paabo recupera DNA da mummie egizie. Fece un clonaggio molecolare (senti)

Il DNA che aveva recuperato non era il DNA della mummia ma il suo.

Quindi c'è tutto un problema legato all'autenticità del DNA.

DNA antico (aDNA)

Con il termine DNA antico (aDNA= *ancient DNA*) si indica qualsiasi traccia di DNA proveniente da un organismo morto o da parte di esso, o anche DNA estratto da campioni biologici non recenti come il DNA contenuto in una goccia di sangue coagulata, nello sperma, o nelle poche cellule epiteliali che si possono ritrovare nel mozzicone di una sigaretta. Quindi se vogliamo essere più precisi si può considerare antico qualsiasi DNA che è stato sottoposto a processi autolitici (sottoposto a carattere enzimatico) o diagenetici (di carattere fisico).

Asse ascisse: anni

Asse ordinate: numero di specie analizzate

Nel corso degli anni sono aumentati i numeri di specie analizzate

Guardando l'anno dall'anno 2004- 2019 → sono aumentate le metodologie molecolari per l'analisi sui genomi e quindi fare studi approfonditi non tanto per il numero di specie ma per le informazioni.

Studi fatti su *Homo sapiens* sono stati fatti utilizzando piccoli frammenti di DNA mitocondriale.

Hanno utilizzato la metodologia classica: è una metodologia che prevede questi step:

estrazione del DNA → quantificazione DNA (real Time-PCR) → amplificazione mediante PCR → clonaggio (perché si tratta di DNA altamente degradato) → sequenziamento del DNA clonato

Oggi la metodologia classica non può usata, ma si usa il sequenziamento NGS.

Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans

David Caramelli¹, Carles Lalueza-Fox², Cristiano Vernesi³, Martina Lari⁴, Antonella Casoli⁵, Francesco Mallegni¹, Brunetto Chiarelli¹, Isabelle Dupanloup⁶, Jaume Bertranpett⁷, Guido Barbujani⁸, and Giorgio Bertorelle^{1,11}

LAST FEWS YEARS

PLoS BIOLOGY

No Evidence of Neandertal mtDNA Contribution to Early Modern Humans

David Serre¹, André Langaney^{2,3}, Mario Choch⁴, Maria Teschler-Nicola⁵, Maja Paunovic⁶, Philippe Menecier⁷, Michael Hofreiter⁸, Göran Possnert⁹, Svante Pääbo¹⁰

¹ Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ² Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris, France; ³ Laboratoire de Génétique et Biologie, Université de Genève, Genève, Switzerland; ⁴ Department of Anthropology, Natural History Museum, Vienna, Austria; ⁵ Institute of Quaternary Paleontology and Geology, Catalan Academy of Sciences and Arts, Zagreb, Croatia; ⁶ Angstrom Laboratory, Uppsala University, Uppsala, Sweden

Neandertal Evolutionary Genetics: Mitochondrial DNA Data from the Iberian Peninsula

Carles Lalueza-Fox¹, María Lourdes Sampietro², David Caramelli³, Yvonne Pader^{4,5}, Martina Lari⁶, Francesc Calafell⁷, Cayetano Martínez-Maza⁸, Markus Bastir⁹, Javier Fortea¹⁰, Marco de la Rúa¹¹, Jaume Bertranpett¹², and Antonio Rosas¹³

¹Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; ²Laboratori de Antropologia, Departament de Biologia Animal e Genètica, Universitat de les Illes Balears, Palma de Maiorca, Spain; ³Departamento de Paleontología, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain; and ⁴Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain

DNA mitocondriale di 24 mila anni fa senti

Ricostruzione di un frammento di DNA mitocondriale di un neandertaliano rispetto ad un mitocondrio sapiens.

Senti

Con lo studio del DNA mitocondriale (da piccoli frammenti) è stato anche scoperto che i neandertaliani non erano presenti solo in europa ma anche in asia.

Abbiamo focalizzato l'attenzione estraendo DNA solo da frammenti ossei

Pochi anni fa è stato scoperto che un'altra fonte di DNA sono frammenti di capelli.

DNA nucleare

Anche con la metodologia classica è stato investigato il DNA nucleare.

Articolo: secondo questo studio ha dimostrato che il colore del mantello del mammut poteva essere polimorfico: potevano avere sia un mantello marrone, sia rossastro (?)

Se era possibile ciò su individui vissuti nel passato: è stato fatto anche per neandertaliani

Utilizzando metodologia classica è stato analizzato il gene FOXP2: per linguaggio articolato → lo stesso anche nei sapiens. Risenti

2007 su scienze andando ad analizzare il dna nucleare di neandertaliani uno vissuto in italia,

questi due avevano una mutazione su questo gene: ci mise di capire che avevano capelli rossi e pelle chiara.

La mutazione puntiforme sul gene MC1R è una mutazione presente su Homo sapiens ma in un punto diverso → ma arrivati alla stessa caratteristica.

LEZIONE 12.10

A DRAFT SEQUENCE OF THE NEANDERTAL GENOME:

Bozza del genoma neandertal fu prodotta nel 2010. Da questa bozza iniziammo a capire che c'è molto Neandertal in noi dall'1-4% di genoma di Neandertal in noi grazie alla comparazione del genoma dell'uomo di Neandertal con Sapiens.

La bozza della sequenza del genoma di N è stato fatto tramite sequenziamento NGS che ci hanno permesso di andare a leggere in profondità il genoma.

Il genoma è stato ottenuto dall'analisi di più frammenti neandertaliani che sono stati assemblati

Anno 2011: GENETIC HISTORY OF AN ARCHAIC HOMININ GROUP OF DENISOVA CAVE IN SIBERIA: utilizzando sempre le nuove tecniche è stato scoperto un nuovo DNA (di frammenti umani ce ne sono pochissime) Denisova: scoperta eccezionale! Perché per la prima volta si ha la possibilità attraverso analisi genomiche di osservare l'esistenza di un nuovo gruppo umano.

Sono stati trovati solo pochi frammenti ossei: dente e un frammento di falange. Si sono accorte che le sequenze mitocondriali e nucleari appartenevano ad una forma umana mai conosciuta.

Il mitocondrio, come il genoma era totalmente differente dai neandertaliani e sapiens

Un aspetto interessante delle popolazioni Denisoviane avviene grazie al sequenziamento dallo studio del genoma nucleare di questo individuo da uno studio fatto nel 2012 A high-Coverage Genome sequence from an Archaic Denisovan Individual.

Gli aspetti interessanti del genoma nucleare:

Anche la donna di denisova condivideva con Homo sapiens dall'1 al 6% di genoma. Questa frequenza così elevata (6%) era a carico delle popolazioni di Papua Nuova Guinea (Oceania).

Questo ci deve far riflettere su come si possano essere sviluppate le nuove migrazioni umane di Homo sapiens una volta uscito dall'Africa, dove nel continente asiatico ha incontrato Denisova con la quale si è mescolato.

Attraverso l'analisi del DNA mitocondriale si pensava che Neandertal e Homo sapiens non si fossero mai incrociati. È stato anche scoperto che non solo queste differenze a livello genetico sono poche ma addirittura che si sono incrociati.

Se si sono incrociati e in noi è presente porzioni del loro genoma, significa che in un certo senso si sono sviluppati degli ibridi che erano fertili e se erano fertili, la barriera specifica viene meno.

Ne tratterà in seguito: aspetti che riguardano le differenti specie riferite al genere Homo oppure dobbiamo considerare il genere Homo costituito da varie subpopolazioni con caratteristiche

peculiari tra loro, ma che potevano scambiare tra loro geni, quindi potevano essere considerati un'unica specie → forte dibattito

Denisova nuova forma umana caratterizzata solo dal punto di vista molecolare. Ad oggi non ci sono reperti scheletrici sia craniali, sia postcraniali che possono descrivere questa forma umana.

THE COMPLETE GENOME SEQUENCE OF A NEANDERTHAL FROM THE ALTAI MOUNTAINS. Gli studi si approfondiscono: viene analizzato da un unico individuo neanderteliano l'intero genoma dei monti Altai (siberia). L'intero genoma conferma le informazioni che sono derivate dall'analisi del genoma deli 5 neandertaliani di cui ha parlato prima. I Monti altai (siberia) hanno un clima freddo → È un aspetto importante per quanto concerne la conservazione del genoma.

Infatti sono stati ottenuti ottimi risultati dall'analisi di questo individuo.

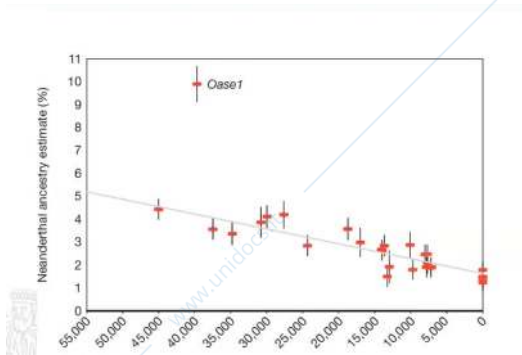
Anno 2014: **A MITOCHONDRIAL GENOME SEQUENCE OF A HOMININ FROM SIMA DE LOS HUESOS.** Nuovo lavoro che riguarda il genoma mitocondriale di un nome rinvenuto a sima de los Huesos (Spagna). La forma umana ritrovata è stata attribuita a Homo heidelbergensis → specie umana afferente al genere Homo da cui si pensa che si siano evoluti sia l'HOMO di neandertal ecc.. Questo individuo aveva un'età di circa 400 mila anni → molto antico. Significa che con le metodologie molecolari a disposizione oggi riusciamo a sequenziare genomi di individui vissuti anche 400 mila anni fa.

Dopo due anni: **NUCLEAR DNA SEQUENCE FORM THE MIDDLE PLEISTOCENE SIMA DE LOS HUESOS HOMONINS:** viene analizzato anche il DNA nucleare di Sima de Los Huesos (per la prima volta). Quindi ad oggi conosciamo genomi mitocondriali e nucleari di tre forme afferenti al genere Homo: Homo neandertalinsis, Homo sapiens, Homo heidelbergensis.

Anno 2016 **THE GENETIC HISTORY OF ICE AGE EUROPE:** studio che riguarda l'analisi molecolare di genomi completi di circa 50 individui che vanno dal paleolitico superiore a circa 30000 anni fa fino ai 5 mila anni fa: vanno a ripercorrere la storia genetica dell'europa del ghiaccio. È interessante perché ponendo in grafico

Asse X: anni asse Y: stima di quanto di neandertaliano c'è in questi individui → la quantità di DNA neandertaliano diminuisce in maniera quasi costante fino ai giorni nostri.

È interessante da un punto da un punta di vista evolutivo in quanto ci siamo accoppiati con loro e il genoma dei nenandertaliani ci forniva dei vantaggi quando ci siamo accoppiati, ma via via che le nostre caratteristiche hanno visto dei cambiamenti climatici importanti, col passare delle generazioni abbiamo eliminato attraverso la selezione il genoma di neandertal dal nostro genoma.



Oase 1 (45 mila anni fa) cade al di fuori della quantità di genoma neandertaliano → studiato nel 2004-2005. Caratteristiche morfologiche particolari. Incrocio tra neandertal e sapiens → è stato stimato quanti generazioni prima della sua c'è stato l'incrocio. Dalla stima è stato visto che uno dei suoi trisnonni era un neandertaliano. 4 generazioni fa c'è stato un incrocio. Articolo uscito nel 2018. Quindi anche se ha intorno a 45 mila anni fa cade al di fuori della quantità di genoma neandertaliano di individui della stessa età: perché il suo trisnonno era neandertaliano. DNA neandertaliano in noi: dall'1 al 4%.

Articolo 2018: RECONSTRUCTING THE GENETIC HISTORY OF LATE NEANDERTHALS:

Ha cercato di ricostruire la storia genetica degli ultimi neandertaliani: Gli ultimi neandertaliani erano un pochino più differenti dai neandertaliani antichi. Ne parlerà alla fine del corso.

NEANDERTHAL BEHAVIOUR, DIET, AND DISEASE INFERRED FROM ANCIENT DNA IN DENTAL CALCULUS. Non si estrae DNA da ossa, capelli, ma anche dal tartaro (dental calculus)

È stato prelevato DNA estratto dal tartaro presenti in denti di alcuni neandertaliani: è interessante perché si riesce a capire il comportamento in quanto nel tartaro si ritrovano molte informazioni della vita quotidiana di un individuo. Ad esempio se un individuo mangia molta carne, si ritroverà nel dental calculus il DNA delle carni ecc...oppure utilizza particolari sostanze per curarsi. Questo perché il tartaro forma una placca dove sotto si conservano tutte le informazioni → Apre scenari interessanti sul comportamento, tipologia alimentazione e malattie del cavo orale.

È stato scoperto che: I neandertaliani che vivevano nel centro Europa mangiavano molta carne.

I neandertaliani che vivevano in ambienti di foresta si nutrivano di frutti della foresta (funghi, pinoli ecc..)

È stato trovato in un Neandertal che presentava un forte ascesso uno dei due mascellari → estirpando il dental calculus → conteneva molecole di pioppo, che contiene acido acetilsalicilico. Quindi i neandertaliani si curavano. Ma quindi se questi avevano la capacità di provvedere alle loro condizioni di salute → il loro psichismo era abbastanza elevato di quanto ci immaginassimo.

Questo è stato ottenuto tramite metodi NGS ed è stato possibile perché si lavavano pochi i denti.

Articolo: A high-coverage Neanderthal genome from Vindija Cave in Croatia: nuove informazioni sui neandertal più recenti: dove si vede quanto neandertal c'è in noi in tutte le popolazioni del mondo: popolazione africana non ha DNA neandertaliano, popolazioni europee e asiatiche ce l'hanno, in modo diverso.

2018-2019: The genome of the offspring of a neandertal mother and a Denisovan father:

Sono stati trovati dei frammenti ossei di una ragazza sui Monti Altai da cui è stato estratto il DNA e ci siamo accorti che era la figlia di un neandertaliano e Denisoviano.

Ma tutto ciò non è affatto semplice, la degradazione e quindi le caratteristiche del DNA antico o degradato condizionano le ricerche. Altro fattore è la contaminazione.

Fattori da tenere in considerazione che possono portare a falsi risultati nell'interpretazione dei dati.

- Degradazione del DNA antico
- Contaminazione

METODOLOGIE per lo studio di genomi antichi.

È diverso rispetto allo studio dei genomi moderni, in quanto ci sono due fattori che influiscono: degradazione e contaminazione.

Non è detto che un DNA più antico di un altro DNA sia più degradato → dipende il DNA dove è stato conservato

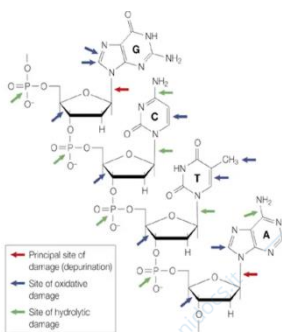
PROCESSI AUTOLITICI E DIAGENETICI

Caratteristiche DNA antico (degradato)

- **Comparsa siti abasici:** le basi azotate della molecola si staccano → se si staccano, DNA si frantuma, quindi di difficile interpretazione. Si osserva frammentazione (da 50-80 ca bp). Se dovessi leggere le 360 bp che formano la regione HV1 del DNA mitocondriale e DNA che recupero è frammentato di circa 1200 bp ciascuna → devo fare almeno 3 letture. Ma se consideriamo un genoma di 3 miliardi di nucleotidi e ho un DNA frammentato di 50-60 bp → si devono fare miliardi di letture.

Siccome la metodologia classica è limitata a leggere solo piccoli frammenti di DNA mitocondriale, nucleare. L'unico modo per leggere i genomi completi si devono usare le tecniche NGS, in quanto il DNA è altamente frammentato

- **Modificazione dei residui glicosidici e di pirimidine:** ho modificazioni riguardo le basi azotate → posso incorrere a false letture o sbagliate della sequenza del DNA. Se una Adenina si trasforma in Iloxantina, questa è letta diversamente dall'adenina, quindi ho una lettura sbagliata. Se la citosina si deamina e viene letta come un uracile, la DNA pol mi copia anziché della Guanina che è complementare ad una citosina, mi diventa una Timina. (vedrà nel dettaglio).
- **Comparsa di cross-link intermolecolari:** cioè di legami tra differenti molecole: ad esempio tra una catena di DNA e proteine formano composti di Maillard → rende il composto illeggibile e inutilizzabile.



Punti in cui avvengono i principali danni al DNA.

Sito di depurinazione: Adenina, guanina dovuti a principali fenomeni ambientali

Siti di danno ossidativo

Siti di danno idrolitico

I danni di carattere idrolitico e ossidativo sono i principali fenomeni ambientali che portano ai processi autolitici e diagenetici.

FATTORI CHE DETERMINANO LA DEGRADAZIONE DEL DNA:

- attività enzimatica cellulare: avviene appena un individuo muore
- attività microbica: batteri funghi presenti all'interno del terreno.
- Ph alti e bassi
- Ossigeno: danno di tipo ossidativo
- Temperatura: alte temperature provocano la denaturazione del DNA
- radiazioni u.v: sterilizza: ad una lunghezza d'onda di 254 nm → dimeri di pirimidine nel DNA
- umidità: danno di tipo idrolitico
- stress meccanico: ossa sotto 10 metri di terra dovuto alla pressione meccanica.

CONTAMINAZIONE: uno dei principali problemi dello studio del Dna antico/degradato umano è la contaminazione che può avvenire da parte dell'operatore o da altro DNA umano amplificato presente sotto forma di aerosol.

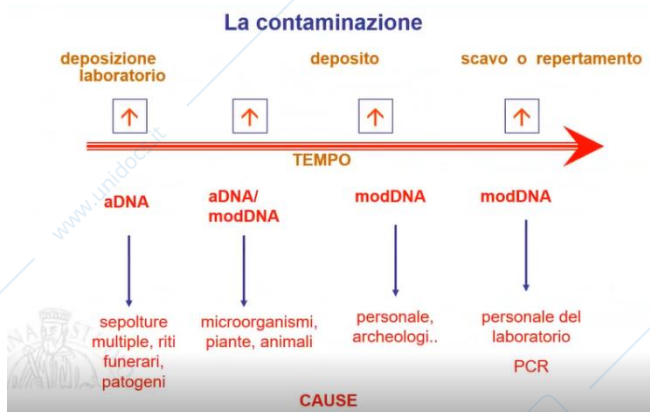
Come possiamo essere sicuri di ottenere il DNA antico autentico?

Ci sono una serie di protocolli che ci permettono di capire quando siamo di fronte a DNA esogeno e DNA endogeno.

Con la metodologia classica ci sono casi ben specifici in cui può essere utilizzata e in altri casi no.

Con la tecnologia NGS si ha la sicurezza grazie ad una serie di procedure sia molecolari sia bioinformatiche che ci permettono di sapere se un campione è sottoposto a fenomeno di contaminazione.

Contaminazione può avvenire durante:

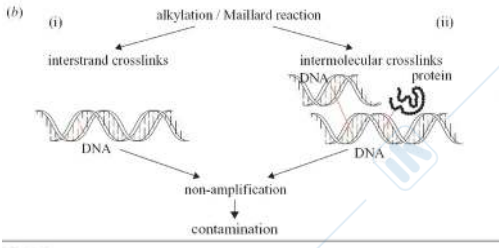


Deposizione in laboratorio, durante il deposito, durante lo scavo o repertamento.

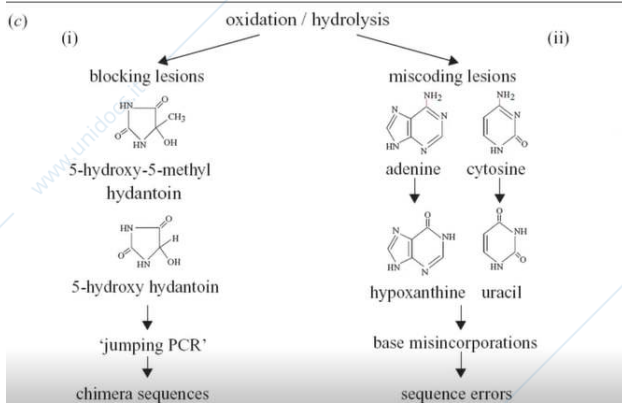
Può essere a carico di DNA moderno di piante, animali, microrganismi, personale, archeologi, antropologo, medico legale, personale del laboratorio.

Danno idrolitico causa la riduzione del numero di molecole che posso recuperare da un'estrazione di un individuo che voglio analizzare e molecole molto frammentate → può portare a fenomeni di contaminazione e quando faccio amplificare non posso amplificare lunghi frammenti perché le molecole di DNA sono corte.

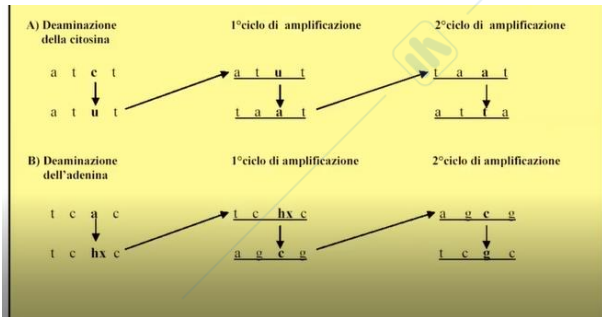
I cross link portano alla formazione dei prodotti di Maillard → provocano una mancata amplificazione del tratto con PCR. Quindi non riesco ad analizzare il DNA



Miscoding lesion: Sono lesioni che portano a letture errate di basi del DNA.



Quindi reperti che sono stati sottoposti a idrolisi spinta (inzuppati in acqua) → Adenina si può trasformare in ipoxantina, e citosina in uracile → ho un'errata incorporazione della base durante la reazione della PCR.



Se sulla catena del DNA ho una citosina che si è deaminata in U, la DNA pol incorporerà nel 2° ciclo di replicazione la T.

Quindi determina una misincorporazione → C verso T

Nel caso della deaminazione dell'adenina: Ipoxantina. La Dna pol nel primo giro di amplificazione incorpora la citosina, e nel 2° giro la guanina. Quindi la misincorporazione sarà. A → G

TABLE 1 Overview over different types of damage in ancient DNA

Type of damage	Process	Effects on DNA	Possible solutions	
Strand breaks	Degradation by microorganisms	Reduction of overall DNA amounts	PCR of short length and overlapping fragments	
	Nucleases in the postmortem cell	Size reduction		
	Other chemical processes			
Oxidative lesions	Damage to bases	Base fragmentation	PCR of short length and overlapping fragments	
	Damage to deoxyribose residues	Sugar fragmentation		
		Nucleotide modification		
DNA crosslinks	Reactions between DNA as well as DNA and other biomolecules	e.g., Maillard products	PTB (N-phenylthiozofium bromide)	
Hydrolytic lesions	Loss of amino groups	Change of coding potential	Multiple independent PCRs Cloning and sequencing of several clones	
				1. adenine ⇒ hypoxanthine
				2. cytosine ⇒ uracil
				3. 5-methyl-cytosine ⇒ thymine
				4. guanine ⇒ xanthine

Si andrà a vedere a seconda del tipo di danno quali sono i processi, gli effetti e le soluzioni.

Dopodiché si costruisce il flusso di lavoro che si fa con la metodologia classica.

TUTTO QUESTO CI PERMETTE ALLA FINE DI ANDARE A RICOSTRUIRE LA STORIA EVOLUTIVA DELLA SPECIE.

Lezione 14.10 5 lezione

La scorsa volta abbiamo trattato di quali possono essere le possibili soluzioni quando ci troviamo di fronte a molecole di DNA degradato. Queste possibili soluzioni le stiamo sviluppando nel Workflow della metodologia classica. Con tale metodologia non possiamo analizzare grossi frammenti di DNA, ma piccoli frammenti.

Per cui oggi parleremo di come si ricostruisce un frammento di DNA mitocondriale estratto da un reperto neandertaliano.

Ricorda: anche se la metodologia classica in campo antropologico è ormai superata, la si può utilizzare anche in campo forense e in alcuni casi può essere sviluppata in campo antropologico, ma non solo, gli studi fatti fino al 2007-2008 sono stati fatti con questa tipologia.

TABLE 1 Overview over different types of damage in ancient DNA

Type of damage	Process	Effects on DNA	Possible solutions
Strand breaks	Degradation by microorganisms Nucleases in the postmortem cell Other chemical processes	Reduction of overall DNA amounts Size reduction	PCR of short length and overlapping fragments
Oxidative lesions	Damage to bases	Base fragmentation	PCR of short length and overlapping fragments
	Damage to deoxyribose residues	Sugar fragmentation Nucleotide modification	Multiple independent PCRs Cloning and sequencing of several clones
DNA crosslinks	Reactions between DNA as well as DNA and other biomolecules	e.g., Maillard products	PTB (N-phenylacetyl thiazolium bromide)
Hydrolytic lesions	Loss of amino groups 1. adenine ⇒ hypoxanthine 2. cytosine ⇒ uracil 3. 5-methyl-cytosine ⇒ thymine 4. guanine ⇒ xanthine	Change of coding potential	Multiple independent PCRs Cloning and sequencing of several clones

Prima riga: AMPLIFICAZIONE DI PICCOLI FRAMMENTI CHE TRA LORO SI SOVRAPPONGONO

È una tecnica che permette di selezionare e amplificare il DNA, ideata da “Malles” intorno al 1983-84 consiste:

- 1) Denaturare DNA
- 2) Aggiunta alla miscela di reazione la coppia di primers che vanno a selezionare il

frammento da amplificare. La coppia di primer si lega al frammento da amplificare ad una temperatura che va da 50-60 gradi che dipende dalla lunghezza dei primer, dalla composizione in C-G

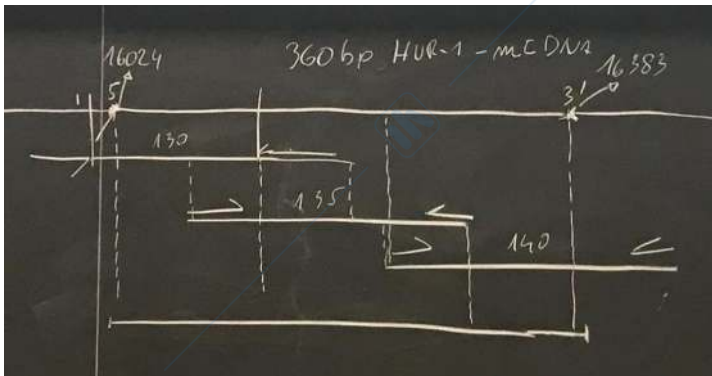
3) Fase di estensione: Taq polimerasi è una polimerasi, proveniente dall'organismo termofilo *Thermus aquaticus*, un batterio che vive ad alte temperature: inserire i dNTP complementari alla catena di DNA da amplificare. La temperatura di estensione è intorno a 72 gradi, in cui la catena viene estesa in entrambe le direzioni

Fino al 3 ciclo non abbiamo la copia target di DNA che vogliamo amplificare. Dal 3 ciclo in poi abbiamo la copia target di DNA da amplificare → a questo punto si crea una situazione a carattere esponenziale per cui da una molecola di DNA si raggiungono più o meno 1 miliardo e 73 milioni di molecole “teoricamente” tutte uguali tra loro.

“TEORICAMENTE” → quando si tratta di analizzare DNA umano moderno → alta probabilità di ottenere molecole uguali tra di loro.

Se si tratta di analizzare DNA antico posso avere il rischio di avere gruppi di molecole uguali tra di loro, ma non tutte molecole uguali tra di loro, perché il prodotto di amplificazione può iniziare da una molecola o da un'altra a seconda di dove si legano i primers.

Fare una PCR con frammenti che fra loro si sovrappongono → si utilizzano più coppie di primer che vanno ad amplificare un frammento (per semplicità non ha disegnato il frammento complementare)



Ho un frammento di 360 bp HVR1 di DNA mitocondriale (16024-16383) umano.

Per amplificare queste 360 bp con una sola coppia di primer (Forward e Reverse) → perché il frammento di DNA di 360 bp li troverò frammentate → si presentano in frammenti da 50-120 bp.

Quindi amplificare 360 bp con un colpo di

amplificazione è impossibile! Infatti se avessimo un prodotto di amplificazione di 360 bp su un campione antico → è probabilmente un contaminante → dell'operatore.

Quindi utilizzare PCR of short length and overlapping fragments significa costruire un sistema ingegnerizzato geneticamente che mi va a fare amplificazione di frammenti più corti.

Costruisco quindi più coppie di primer che amplificano frammenti diversi (come sulla lavagna): in questo esempio utilizzo 3 coppie di primer per questo frammento lungo 360 bp.

RICOSTRUZIONE SEQUENZA CONSENSO → Confrontare la sequenza dei primer e dei prodotti di amplificazione che ho ottenuto per ricostruire la mia sequenza consenso grazie alla presenza di zone uguali nella zona di sovrapposizione.

Ci sono metodi bioinformatici per allineare le estremità e quindi ricostruire la sequenza di DNA.

Questo aspetto è importante non solo nell'aspetto antropologico per lo studio della storia naturale del genere Homo, ma anche in ambito forense → perché se sulla scena del crimine ho DNA degradato → voglio amplificare il DNA mit con la metodologia classica → un modo è utilizzare questa tecnica.

LESIONI DI CARATTERE OSSIDATIVO

- danneggiano le basi → frammentazione delle basi
- danneggiano i residui di deossiribosio → frammentazione zuccheri
- creano misincorporazione di nucleotidi

Per i primi due punti → utilizzo la tecnica precedente

Per il 3 punto → Soluzione multiple e indipendenti PCR, clonaggio e sequenziamento di molti cloni.

- ➔ Se devo fare PCR multiple → devo fare multipli estratti → più estratti dallo stesso individuo → quindi devo ripetere più volte l'esperimento di PCR of short length and overlapping fragment.

In passato estraevano il DNA da frammenti di ossa lunga, frammenti di femore, coste (chiamati anche post-craniali) e qualche dente → risultati abbastanza buoni ma non eccellenti come le recenti estrazioni da altri distretti ossei.

Utilizzavano questi frammenti prima perché gli archeologi e panteologici che recuperavano questi frammenti davano più volentieri da analizzare una costa piuttosto che un dente → quindi si adattavano.

Dopo le PCR multiple → si passa al clonaggio (vedi slide): clonaggio dei prodotti di amplificazione →

DEFINIZIONE CLONAGGIO: è un'amplificazione in vivo rispetto alla PCR che è in vitro.

Con il termine clonaggio si intende l'inserimento di frammenti di DNA inseriti in vettori (plasmidi) nel multicloning site. A questo punto si ha la trasformazione del plasmide nella cellula competente (la competenza è la capacità di una cellula di acquisire DNA estraneo dall'ambiente) di E.coli → viene seminata su un terreno di coltura → le cellule moltiplicandosi, moltiplicano al loro interno anche il plasmide con l'inserito (il frammento di DNA).

COME SI DISTINGUONO LE CELLULE CHE HANNO INCORPORATO IL PLASMIDE CON INSERITO? Le cellule competenti vengono fatte crescere su un terreno di tipo selettivo:

- Le cellule che hanno incorporato il plasmide con l'inserito a livello morfologico sono colonie bianche → perché il gene viene posizionato all'interno del gene LacZ → quindi il gene viene interrotto → causa una non funzionalità del gene LacZ → quindi non si ha codifica della beta galattosidasi → non degrada lattosio → colonie bianche

* LACZ codifica β-galattosidasi è un enzima idrolitico che catalizza la scissione del lattosio in glucosio e galattosio

Le cellule che hanno incorporato il plasmide ma NON l'inserito → il gene lacZ è funzionante → codifica l'enzima β galattosidasi → degrada il lattosio → colonie diventano Blu → quindi sono da scartare

Le cellule che NON hanno incorporato il plasmide → muoiono perché sono in un terreno selettivo contenente ampicillina/kanamicina (perché il plasmide ha i geni che conferiscono resistenza all'antibiotico)

Perché si utilizza il clonaggio? E cosa vado a clonare?

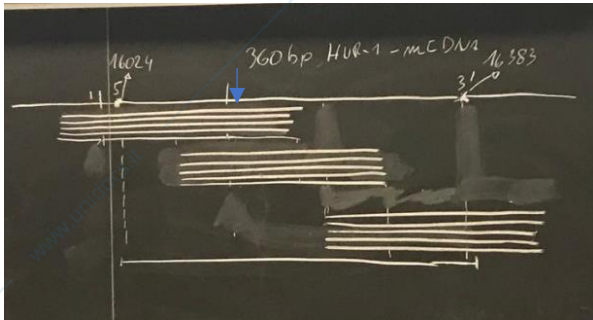
Vado a clonare i prodotti di amplificazione: si parte da un DNA che è degradato (quindi frammentazione e misincorporazioni) → quindi all'interno della miscela di DNA avrò molecole non degradate e molecole degradate, molecole endogene e probabilmente anche molecole esogene → quindi se utilizzo una coppia di primer → si attacca dove trova complementarietà → se si attacca su un frammento che ha misincorporazioni o si attacca su un frammento esogeno (più probabile) → il prodotto di amplificazione che contiene DNA esogeno, DNA endogeno, DNA endogeno degradato.

Se li sequenzio (con Sanger modificato, senza fare clonaggio → siccome all'interno del DNA che ho estratto ci possono essere 3 gruppi di prodotti di amplificazione → sequenziando direttamente un prodotto di amplificazione con le caratteristiche di DNA degradato NON so quale gruppo di molecole sequenzio → propaga l'errore. Quale DNA sequenzio di questi 3 gruppi? Può essere quello in maggiore quantità ecc..

Utilizzando il clonaggio → separo i gruppi di prodotti di amplificazione → ciascuna cellula competente mi riproduce un frammento → sequenzio il prodotto del clonaggio che si chiama amplicone →

Quindi il clonaggio del work flow serve come SETACCIO per separare gruppi di molecole generati dall'amplificazione del DNA

Quindi se da ciascun frammento ho più frammenti che mi provengono dal clonaggio ho questa situazione a seconda del numero di ampliconi che avrò sequenziato per ciascun prodotto di amplificazione



COME FACCIAMO A CAPIRE TRA QUEL GRUPPO DI AMPLICONI CHE HO SEQUENZIATO QUALI SONO QUELLI ESOGENI, ENDOGENI ECC?

Perché vado a considerare la mia sequenza nucleotidica → Si suppone in questa posizione qualcosa che nella sequenza di riferimento (che è

depositata nella banca dati a cui metto a confronto le sequenze che ottengo)

Una volta che ho sequenziato → le mie sequenze le confronto con una sequenza di riferimento che è depositata nella banca dati. Si suppone che nella sequenza di riferimento in posizione indicata dalla freccia ci sia una C → nella mia sequenza trovo tutte T in quella posizione. Lasciando così posso dire che sono stato sfortunato in quanto separando i miei ampliconi ho trovato tutte molecole che possedevano una mutazione o una misincorporazione. Ma posso vedere nel secondo frammento, stessa posizione che si sovrappone al primo → se mi trovo tutte T → allora la probabilità che in quel punto anziché una misincorporazione, ho una sostituzione nucleotidica è più alta → perché se prendo un'altra coppia di primer e faccio un'altra amplificazione di frammenti differenti che si sovrappongono in parte al primo → ottengo nello stesso punto ho la stessa modificazione nucleotidica.

Siccome si fanno multiple e indipendenti PCR → si prende un altro estratto di un differente distretto osseo proveniente dallo stesso campione e rifaccio PCR e clonaggio → se dal secondo estratto ottengo la stessa sostituzione nucleotidica → si abbassa la possibilità che la sostituzione sia dovuta ad un errore di misincorporazione.

Esistono una serie di criteri che vengono utilizzati per i studi sul DNA degradato e vedere se c'è uno stesso risultato.

Lesioni idrolitiche → porta, come i danni ossidativi a misincorporazioni → quindi si procede come i danni ossidativi

APPLICAZIONE PRATICA - DETERMINAZIONE DEL SESSO VIA MOLECOLARE mediante METODOLOGIA CLASSICA

Si fa la determinazione del sesso su resti scheletrici. Perché? Per due motivi

i) Dal reperto scheletrico → Bacino → differente tra un uomo e donne → perché ha il canale del parto. Lo si può riconoscere anche dal cranio, ma è più difficile. Lo si determina in individui adulti abbastanza conservati.

Nei bambini (infant e subadulti) è impossibile da un punto di vista morfologico e morfometrico determinare il sesso.

La determinazione del sesso in individui antichi (ma anche moderno, anche se si vede) → si amplifica un frammento del gene → gene omologo della amelogenina (che fa parte dello smalto dentario) che ha un polimorfismo di lunghezza sui cromosomi sessuali (lunghezze differenti su cromosoma X e Y). Pertanto il gene omologo della amelogenina che sta sullo stesso locus su differenti cromosomi ha lunghezze differenti su cromosoma X e su cromosoma Y.

→ 112 bp sul cromosoma Y

→ 106 bp sul cromosoma X

Profilo maschio: 112, 106.

Estrazione → PCR → visualizzazione del prodotto di amplificazione su gel di agarosio → corsa del prodotto di amplificazione attraverso l'elettroforesi su gel di agarosio.

Visualizzo nel seguente modo (non ho la foto)

Gel agarosio: Ci sono i pozzetti dove si inserisce il DNA con una micropipetta. Successivamente si attacca alla corrente e il DNA andrà dal polo (-) al (+). La corsa del DNA all'interno di un gel agarosio è inversamente proporzionale alla lunghezza del DNA → tanto è più corto il mio frammento → tanto più corre → quindi ci si aspetta una banda verso il basso circa (verso il polo positivo). Nel caso dei due frammenti di lunghezza differente → ci aspettiamo che il cromosoma Y che è di 112 bp è più lungo → quindi incontra più resistenza nella corsa elettroforetica → si troverà quindi al di sopra della banda del cromosoma X.

pozzetto: K⁻ → controllo negativo: c'è la miscela di reazione ma senza il DNA → quindi non si deve ottenere nessuna banda → se si ottiene una banda si considera la corsa elettroforetica contaminata e si fa di nuovo →

Pozzetto K⁺ → controllo positivo → miscela di reazione più un DNA il cui funzionamento è sicuro → dobbiamo ottenere per forza una banda.

Avvio l'elettroforesi, copro la pista con carta d'alluminio e aspetto circa mezz'oretta.

Guardando i risultati → i controlli erano puliti → quindi non ho contaminato e i miei reagenti non erano contaminati, non ho sbagliato la PCR.

Campione 1 e Campione 2 e campione 3 → frammenti che hanno corso alla stessa lunghezza →

Campione 1: maschio, campione 2 e 3 → femmina.

Il k⁺ è maschio, perché c'ha entrambi gli alleli.

Le deduzioni sono corrette ma c'è una riflessione da fare →

Non torna che nel campione 1 c'è una banda sola → essendoci la banda di 112 basi → e il controllo è pulito → e non c'è il frammento di 106 → deduco che è maschio → può capitare che il frammento di 106 basi era molto degradato, quindi i primer non hanno trovato una zona in cui attaccarsi →

Campione antico che contiene del DNA degradato può capitare che non ci sia l'amplificazione di un frammento.

Questi campioni vengono dal solito contesto di scavo → Sorge il dubbio: i campioni 2 e 3 che mi danno i due campioni di 106, è probabile che anche loro abbiano perso gli alleli di 112 → quindi sono due maschi anche il campione 2 e 3 → possono incorrere ad errori.

Per questo la determinazione del sesso con la metodologia classica non è sufficiente farla solo con il gene omologo della amilogenina ma → bisogna fare una co-amplificazione con il frammento di gene SRY di 93 bp che è caratteristico del cromosoma Y.

Quindi faccio un'amplificazione multiplex → metto due coppie di primer che mi amplificano contemporaneamente il gene omologo della amilogenina e SRY → posso avere un risultato differente

Campione 1: dove era presente l'allele di 112 è presente anche l'allele di 93 bp

Campione 2: sesso è maschio

Campione 3: è femmina

QUINDI NEI CAMPIONI ANTICHI POSSO AVERE PERDITE ALLELICHE → quindi per la determinazione del sesso in campioni antichi → devo fare una PCR multiplex co-amplificando il gene omologo dell'amilogenina con il gene SRY, ma devo ripetere l'esperimento molte volte e ottenere sempre lo stesso risultato.

LEZIONE 21 OTTOBRE

Avevamo parlato a cosa serve il clonaggio. Es. determinazione del sesso in reperti umani antichi.

Le analisi fatte su DNA degradato attraverso il work flow della metodologia classica prevede una serie di regole chiamata "Golden criteria" → sono stati evidenziati per la prima volta da Alan Cooper, pubblicati su rivista *Science: Ancient DNA do it right or not all* (articolo)

Questi criteri sono riportati in questa tabella

criteria of authenticity
physically isolated pre-PCR facility nightly UV irradiation of surfaces, isolated ventilation, flow hoods or glove boxes, daily cleaning with bleach and movement only from pre-PCR to post-PCR buildings/areas
blank controls extraction and PCR controls should be performed at a ratio of at least 1 : 5 and 1 : 1 with samples, respectively; similar ratios are required for culturing experiments
independent reproducibility a subset of results should be verified by independent replication in another laboratory to rule out intra-laboratory contamination; critical for key results
cloning and sequencing a subset of PCR products should be cloned to assess damage, contamination, and to detect nuclear insertions (numts); repeat extractions and amplifications should also be assessed
time-dependent pattern of damage and diversity provides strong support for antiquity in microbial studies
decontamination of reagents and specimens bleach, acid, ultra-filtration, baking and UV irradiation of reagents and tools should be routine; particularly important in ancient human and microbial studies. The surface of samples should be removed.
DNA from associated remains can provide strong circumstantial support in ancient human and microbial studies
appropriate molecular behaviour an inverse relationship between amplification length and strength support authenticity; expected copy number relationship (e.g. between mtDNA, chloroplast DNA (cpDNA) and nuclear DNA (nuDNA))
uracil-N-glycosylase (UNG) treatment UNG removes deamination products of cytosine, and is particularly important in studies where results involve few substitution differences
quantification of starting templates particularly important in studies where interpretation is based on few substitution differences, or few specimens
biochemical preservation measures of diagenesis can indicate whether DNA survival is likely

1. Aree di lavoro fisicamente isolate: si intendono dei laboratori in cui vengono fatti solo particolari tipi di esperimenti: se devo estrarre il DNA da un reperto antico → dovrò avere un lab in cui faccio l'estrazione del DNA (Extraction Room) → in questo locale si portano frammenti biologici (di ossa ecc..) → questi devono essere trattati in un determinato modo

Es. reperto osseo (dente, frammento osseo lungo, rocca pretosa) di cui devo fare l'estrazione → prima di entrare in questa stanza e aprire il mio reperto osseo che viene da uno scavo, da un museo ecc... → devo utilizzare protezione individuale (coprire con camici sterili, mascherine sterili, visiera ecc..) → una serie di precauzioni che mi impediscono di apportare il mio DNA sul reperto che voglio andare ad analizzare

Questa area in cui faccio l'estrazione nella metodologia classica è chiamata → PRE-AREA (Pre PCR area- area prima della PCR) → una stanza dove faccio l'estrazione del mio campione: deve essere fisicamente separata da tutte le stanze in cui utilizzo il DNA che è amplificato.

Caratteristiche della stanza: Ha una cappa chimica; la pressione dell'aria è positiva → viene spinta al di fuori; ha al suo interno meccanismi di sterilizzazione (raggi UV che si accendono quando non si fanno esperimenti)

Quindi avere stanze separate significa suddividere i laboratori in cui si fa analisi su DNA degradato in locali → in ciascun locale si fanno particolari tipi di analisi:

Quindi laboratori in cui si lavora su DNA antico è DIVERSO da laboratorio in cui si lavora su DNA moderno

Quindi ho una stanza in cui si fa l'estrazione e una stanza in cui si allestisce la reazione di PCR → vengono chiamate insieme PRE-PCR AREA

Ci sono stanze chiamate POST-PCR area → dove avvengono le fasi post-PCR: costituita da più stanze

- Stanza dove è presente il termociclatore (macchinari in cui viene fatta la PCR)
- Stanza che viene utilizzata per il clonaggio genico
- Stanza per visualizzazione dei prodotti di amplificazione

- Stanza utilizzata per sequenziare il DNA

Queste separazione riguarda nella metodologia classica.

Quando si parlerà di NGS, anche in questo caso sono separate ma con funzioni differenti.

Quindi ogni stanza servirà per un determinata attività di ricerca e devo seguire un flusso di lavoro che permettere di mantenere alta la concentrazione su eventuali apporti contaminanti.

2. **BLACK CONTROL**: controlli negativi messi sia in fase di estrazione, sia amplificazione per continuare con gli esperimenti. Se il controllo non è negativo → i reagenti utilizzati per l'estrazione sono contaminati.

3. **INDEPENDENT REPRODUCIBILITY** (utilizzato solo nella metodologia classica). Fare un analisi su reperto scheletrico comporta anche la riproducibilità del dato. Quando si fa un esperimento in laboratorio, non è detto che il risultato sia valido, quindi l'esperimento deve essere replicato con la metodologia classica da un altro operatore e in un altro laboratorio → e quindi riproducibilità del dato in un altro laboratorio. I risultati devono essere uguali.

4. **CLONING AND SEQUENCING**: (abbiamo visto prima)

5. **TIME DEPENDENT OF DAMAGE AND DIVERSITY**: alcune volte si possono osservare dei pattern, dei modelli di frammenti di DNA che vanno incontro a particolari misincorporazioni → in particolari ci sono regioni che sono più pronte a subire danni post mortem rispetto ad un altre → marcatore importante per la riproducibilità del danno. Molti hanno utilizzato queste regioni per verificare l'autenticità del dato.

6. **DECONTAMINATION OF REAGENTS AND SPECIMENS**:

Decontaminazione dei reagenti → vengono venduti sempre senza DNA o RNA.

Decontaminare un reperto è interessante:

→ viene pulito → significa liberarlo dallo sporco → viene rimosso lo strato superficiale di 1-2 mm con microtrapani → messo in un macchinario con raggi UV a 250 nm di lunghezza che sterilizza la superficie (ma non esclude che il campione possa avere all'interno contaminante esogeno). Es: Rocca petrosa va divisa in 2 e all'interno si recupera la polvere d'osso che interessa per fare l'estrazione → meno prone a subire contaminazioni dall'esterno. Mentre ossa lunga come femore, ossa lunga ecc.. sono molto prone a subire contaminazioni. I denti invece sono elementi abbastanza importanti per estrarre il DNA anche se si possono fare molte analisi (es. recuperare dental calculus) → elementi anatomici che gli archeologi non danno molto volentieri.

7. **DNA from associated REMAINS**: DNA da reperti associati alla sepoltura: quando un antropologo va a recuperare una sepoltura, in questa sepoltura si ritrovano anche frammenti ossei animali. Nella sepoltura della donna di paglicci molto alta e ben strutturata dove sono stati trovati molti frammenti di ossa animali (denti che formavano una collana o animali morti li successivamente alla donna di paglicci).

Fare il DNA di queste ossa di animali e analizzare il DNA su cui si vuole fare uno studio di carattere evolutivo è un'operazione che nella metodologia classica è molto importante → se utilizzo una coppia di primer che mi vanno ad amplificare il DNA umano e utilizzo questa coppia di primer su un DNA estratto non umano → succede che non ho prodotto di amplificazione. Si suppone che ottengo un prodotto di amplificazione da DNA estratto da ossa di animali → potrebbero essere successe due cose: ho confuso ossa di animali con ossa umane (abbastanza raro),

oppure in ossa di animali è presente del DNA umano (più probabile) → ma se queste ossa sono state recuperate contestualmente allo scheletro completo della donna di paglicci → posso ipotizzare che chi ha fatto il recupero possa aver contaminato con il suo DNA sia il DNA animale sia il DNA dello scheletro umano → Quindi andare ad analizzare reperti associati permette di escludere una contaminazione di “massa” (risenti).

8. **APPROPRIATE MOLECULAR BEHAVIOUR** (appropriato comportamento molecolare) con la metodologia classica recuperare frammenti lunghi 400-500 di basi su un reperto antico non è una cosa che può funzionare perché il DNA è molto antico → quindi utilizzare tutti quei sistemi che abbiamo visto che permettono di essere abbastanza sicuri che ciò che andiamo a cercare sono in grado di trovarlo. Altra considerazione: se trovassi su un reperto antico con la PCR un frammento lungo di circa 500 bp → è un esogeno contaminante.

9. **UNG TRATTAMENT**: rimuove i prodotti di deaminazione della citosina, ma anche dell'adenina → fare un trattamento con UNG significa andare a rimuovere quei frammenti di DNA che contengono misincorporazioni → non si amplificano tali frammenti. Corro il rischio di non avere prodotto di amplificazione. Anche con l'NGS viene utilizzato, ma non per lo stesso motivo.

10. **QUANTIFICATION OF STARTING TEMPLATE:**

Quantificazione DNA estratto → è importante per capire la probabilità di avere risultati negli esperimenti successivi. (con PCR competitiva e poi con RT-PCR).

E' stato stimato che nell'aria sono presenti molte quantità di DNA mitocondriale (da 1000-1500-2000 molecole). Quindi se devo fare un esperimento su un campione antico con la metodologia classica → avrebbe significato che se il frammento da cui estraggo il DNA avesse un numero inferiore alle molecole che ho nell'aria? No, perché molto prob le molecole che vado ad amplificare non sono le molecole del reperto, ma delle molecole con cui il reperto è venuto a contatto.

La quantificazione attraverso PCR real Time (misura la quantità di DNA presente nel campione) → è possibile stimare in tempo reale il numero di molecole presenti in un estratto in base alla quantità di prodotto di amplificazione che ho ottenuto dopo aver sviluppato la PCR real time.) Se quantificazione <1000 molecole → si rifaceva l'esperimento, perché prob contaminante. Se > → si proseguiva

È una quantificazione fatta anche con NGS → per quantificare le concentrazioni delle librerie a DNA.

11. **BIOCHEMICAL PRESERVATION**: in passato si andava a vedere alcune misure che riguardavano la diagenesi delle ossa. Si andavano a vedere parametri che indicano la prob di trovare DNA in buono stato di conservazione: rapporto di “racemizzazione” dell'acido aspartico D sull'acido aspartico L. Le analisi termo-gavimetrica (quanta materia organica era presente sul reperto che è stato recuperato).

Un buon rapporto è intorno su 0,1 → molto L e poco D perché il passaggio da L a D era proporzionale alla degradazione del DNA.

Più il DNA è degradato → il rapporto aumentava. Funziona questo parametro ma non sempre, quindi non si utilizza più.

CONFRONTO DEI CRITERI DI AUTENTICAZIONE che vengono sviluppate nelle scienze forensi e chi si occupa di antropologia molecolare.

Criteria of authenticity

	Forensic genetics	Molecular archaeology
Separated workstation and labware	Yes	Yes
Investigation of biochemical preservation	-	Yes
Clean extract/PCR controls	Yes	Yes
Cloning of PCR products	-	Yes
DNA quantitation	Informative	Yes
Phylogenetic test	-	Yes
Independent reproduction of the results	Yes	Recommended for human remains

Come si vede dalla figura (legge cosa c'è scritto)

→ Come si può vedere le due materie si assomigliano molto.

Qui subentra un punto interessante → **FILOGENETIC TEST** (test di carattere filogenetico)

Phylogenetic test: se devo fare la filogenesi di una popolazione o di un individuo utilizzo le informazioni genomiche e genetiche che provengono dall'analisi di questo individuo, attraverso la costruzione di alberi filogenetici che mettono in correlazione le differenze e somiglianze di due individui o popolazione, attraverso particolari modelli di calcolo. Quindi attraverso questi strumenti guardo a chi è più simile chi è più simile questo individuo rispetto ad un altro mi permette di ricostruire la storia evolutiva di questo individuo confrontandolo con un altro individuo, o di una popolazione confrontandola con un'altra popolazione.

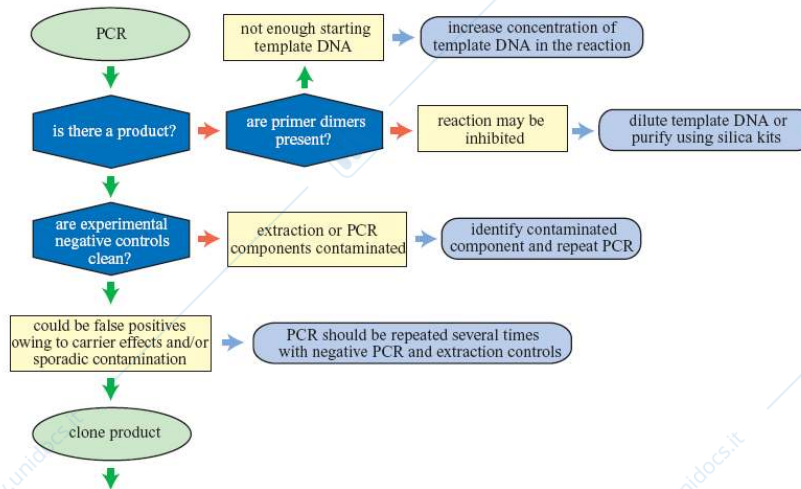
Es. filogenesi delle sequenze mitocondriali di Homo sapiens, si sono separate da quelle di Neandertal circa 800 mila anni fa → andando a ricostruire la storia evolutiva.

Un test di carattere filogenetico è interessante perché ci permette mi stupirei molto se attraverso l'analisi del test filogenetico fatto su un nativo americano, andando a recuperare il DNA mitocondriali di un nativo americano, questo mitocondrio clusterizzasse (cadesse) all'interno della variabilità delle popolazioni europee. So che i nativi americani hanno aplogruppi A, B, C, D e X che è differente da un aplogruppo delle popolazioni europee → posso ipotizzare che tutto ciò che ho fatto possa essere inficiato da forti problemi di contaminazione? Quindi quel DNA non è del nativo americano, ma è il DNA della persona che ha manipolato il nativo americano, recuperando il DNA dell'operatore e non del nativo americano. **Quindi fare test filogenetico ci permette di capire se le analisi che ho fatto hanno un supporto di carattere evolutivo.**

Altro esempio: Mi stupirei molto di trovare in un individuo del paleolitico superiore (25- 20 mila anni fa) un aplogruppo differente dall'aplogruppo U → sappiamo ad oggi (perché ne hanno studiati tanti) → i primi sapiens europei (dai 45-30 mila anni fa o più recente) avevano tutti un mitocondrio con aplogruppo U. Se io in un individuo del paleolitico del superiore trovo un aplogruppo differente da U → due possibilità: contaminazione o scoperta.

FLUSSO DI LAVORO-Work flow con METODOLOGIA CLASSICA

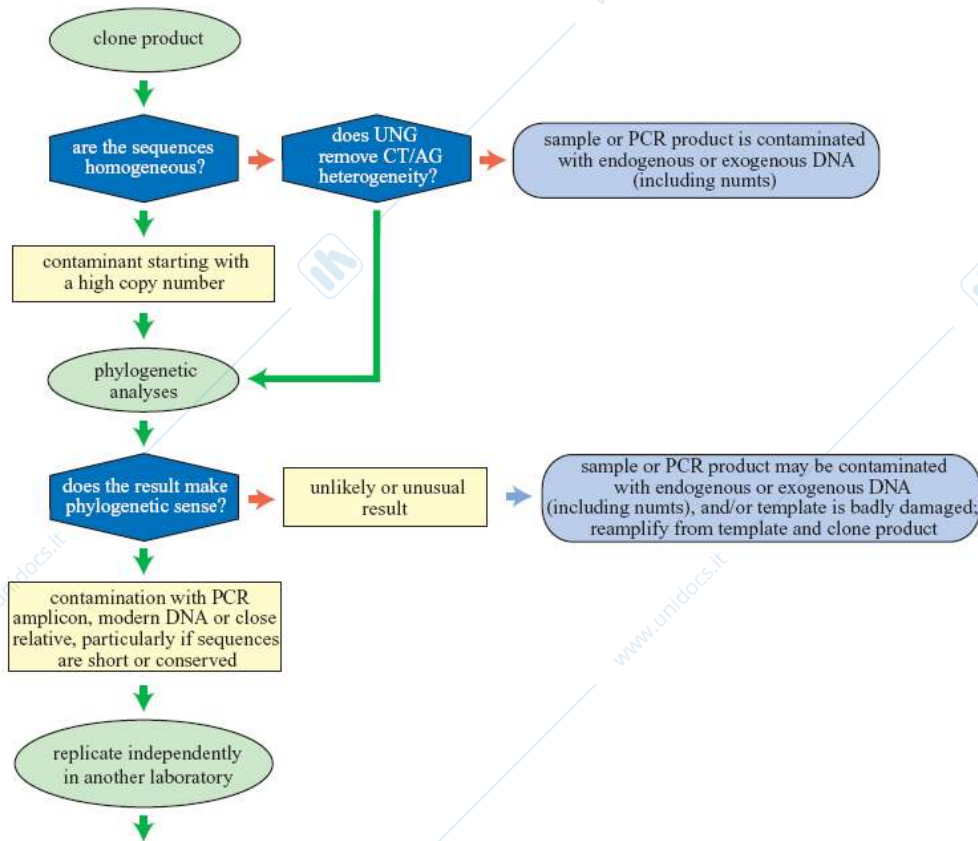
null hypotheses



Si parte da PCR. Seguiamo la freccia verde. C'è prodotto di amplificazione? (lo vedo con l'elettroforesi)

Due possibilità:

- **NO** → Cosa vedo allora??
 - Posso osservare se sono presenti dimeri di primer.
 - **SI** → primer che tra loro si appiccicano perchè trovano tra loro basi complementari → se si autoamplificano → avrò un prodotto di amplificazione di dimeri di primer. Perché lo fanno? Perché non c'è sufficiente DNA di partenza → aumentare concentrazione di DNA.
 - **NO** → (non c'è prodotto di amplificazione e dimeri di primer) → reazione PCR può essere inibita → perché mi son portato avanti nell'esperimento sostanze che mi inibiscono la reazione di PCR → sostanze che sono nel suolo (tannini?) → diluire e purificare il DNA estratto.
- **SI** → son puliti il controllo di amplificazione e estrazione?
 - **NO** → componenti dell'estrazione o PCR sono contaminati → identificare i componenti dell'estrazione o amplificazione e ripetere la PCR, altrimenti buttare tutto e rifare analisi.
 - **SI** → Il mio prodotto di amplificazione potrebbe portare al suo interno dei prodotti falsi positivi o delle contaminazioni di carattere sporadico → **SI!** Perché la PCR su DNA degradato può contenere misincorporazioni, DNA esogeno ecc → si risolve con clonaggio dei prodotti di amplificazione → in questo modo riesco a separare.



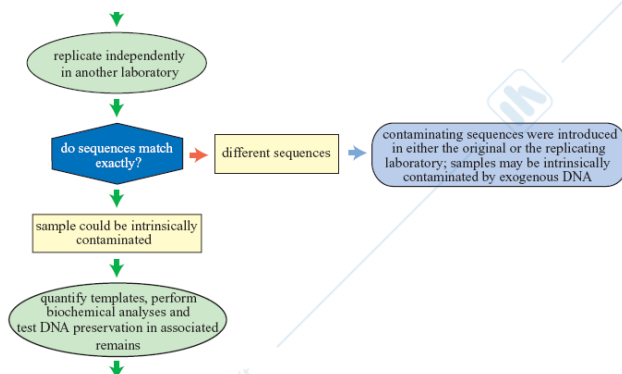
Faccio il Clonaggio → sequenziamento → le guardo

Le sequenze sono tutte uguali tra di loro?

- NO → possono contenere misincorporazioni → utilizzo UNG che mi elimina le misincorporazioni.
- SI → Oppure possono andare avanti perché non ho misincorporazioni o perché sono sparse! Cosa è successo? Ho una contaminazione generalizzata → Faccio il test filogenetico.

I risultati hanno senso filogenetico?

- No (improbabile o risultati inusuali) → il DNA che ho all'interno del reperto è DNA contaminante
- Si → vado avanti



Vado a replicare tutto in un altro laboratorio

Le sequenze combaciano?

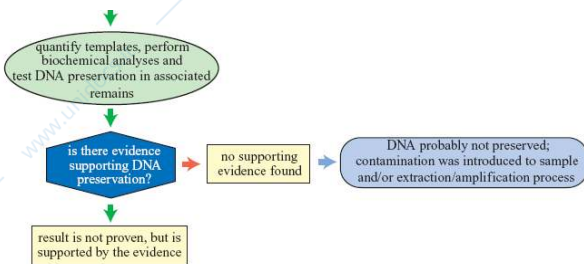
- NO → contaminazione da uno dei due laboratori o magari questo DNA potrebbe essere intrinsecamente contaminato → eliminare studio del campione.

Oppure posso andare avanti: Hanno lo stesso motivo mitocondriale? Vado avanti anche se posso pensare che possa essere intrinsecamente contaminato

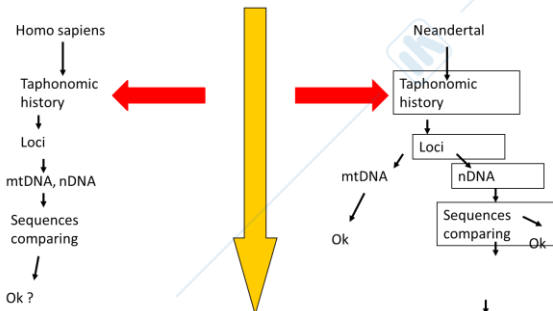
Ad esempio fare il DNA di un individuo etrusco vissuto circa 2000 anni fa nel territorio toscano → DNA mitocondriali molto simili ai nostri → fare il DNA di un individuo etrusco con met classica possono andare incontro a questa situazione, cioè che il DNA è intrinsecamente contaminato e non accorgermene.

Ma posso decidere di andare avanti → Quantificazione template → posso fare test animali associati → verificare cosa succede.

Quello che mi spaventa di più è che anche se faccio tutte queste analisi che seguono questo flusso di lavoro, **in determinati casi il risultato è supportato dalla evidenza scientifica degli esperimenti che ho fatto ma non è totalmente provato → con la metodologia classica posso fare particolari tipi di campioni ma non tutti.**



How to work



Se lavoro con i neandertaliani con la metodologia classico non sbaglio perché hanno un mitocondrio differente dal Sapiens.

È bene conoscere la **storia tafonomica** → ci permette di tracciare la storia di un campione → ci permette di capire chi ha manipolato il campione.

I campioni museali umani ne conosco la storia tafonomica? NO! Quindi non posso studiarli con la metodologia classica (posso conoscere quella degli

ultimi 20 anni, prima no!)

Con la metodologia classica POSSO studiare campioni di cui SI CONOSCE la storia tafonomica →

es. reperto trovato in grotta, facendo il DNA mitocondriale di tutte le persone con cui è venuto a contatto (quindi delle persone che l'hanno manipolato) e il suo mitocondrio è differente dalle persone con cui è venuto a contatto è sicuramente il DNA mitocondriale del reperto.

Quindi se voglio conoscere la storia tafonomica, devono conoscere i loci in cui andare ad investigare:

se vado ad investigare il DNA mit → va bene

se devo andare ad investigare il DNA nucleare di un campione umano antico → non è facile.

RACCONTA ANEDDOTO CHE NON SCRIVO.

LEZIONE 26 OTTOBRE

Due studi che riguardano la metodologia classica:



Available online at www.sciencedirect.com

ScienceDirect

Forensic Science International xxx (2007) xxx–xxx



www.elsevier.com/locate/forensic

Genetic analysis of the skeletal remains attributed to Francesco Petrarca

David Caramelli^{a,*}, Carles Lalueza-Fox^b, Cristian Capelli^c, Martina Lari^a,
 Maria Lourdes Sampietro^d, Elena Gigli^a, Lucio Milani^a, Elena Pilli^a, Silvia Guimarães^a,
 Brunetto Chiarelli^a, Vito Terribile Wien Marin^e, Antonella Casoli^f, Roscoe Stanyon^a,
 Jaume Bertranpetit^b, Guido Barbujani^g

è una ricerca scientifica che collega la parte Antropologia forense e antropologica molecolare. Per far capire come queste due materie sono tra loro interconnesse → questo è un classico studio sviluppato che riguardano la metodologia classica, mostrano tutti gli

steps per recuperare e caratterizzare geneticamente reperti scheletrici attribuibili a Petrarca.

Questo studio nasce perché molto spesso vengono riesumati i corpi di personaggi importanti per vedere lo stato di conservazione e per capire se effettivamente se l'individuo seppellito in quel contesto specifico è l'individuo di cui si parla.

Nel 2005-2006 il prof venne contattato per farlo partecipare a questo tipo di studio → Perché fare questo tipo di analisi, qual è significato scientifico? Perché caratterizzare il genoma di un individuo vissuto molto tempo fa può avere un significato se accanto a questo c'è una domanda di carattere biologico a cui rispondere: ad esempio → il corpo depositato in questa sepoltura appartiene a questo individuo? Per fare un'assunzione di questo tipo occorre avere un genoma di confronto, ad esempio parenti di cui si conosce il DNA per carattere di tipo storico ecc.. Oppure se esistono sussistono delle condizioni di conservazione del DNA che permettono di andare a ricostruire esistono caratteri fenotipici.

Nel caso del prof, c'erano dei dubbi se il corpo fosse tutto di Francesco Petrarca. Perché questo dubbio? Diversi motivi:

1374: La prima sepoltura avvenne all'interno della Chiesa di Arquà

1380: Ci fu una prima verifica della sepoltura e traslazione nell'arca marmorea nel piazzale accanto alla chiesa.

1630: Alla mezzanotte del 27 maggio, frate Tommaso Martinelli da Portogruaro, "spezzata" l'arca all'angolo di mezzodì, trafuga le ossa dell'avambraccio destro del Poeta. Rintracciato dalla giustizia veneziana, venne sottoposto a processo e condannato all'esilio. I reperti trafugati non furono mai ritrovati.

Intorno al 1600 circa è uso comune prendere ossa di una persona importante e rivenderle.

1843: Carlo Leoni, cultore di storia patria, decide di curare il restauro dell'arca e se ne assume l'intera spesa. Nell'occasione, è documentata una rapida ricognizione da parte del prof. Antonio Meneghelli. Viene prelevata anche una costola che viene reinserita nella sepoltura due anni dopo.

1873: Nuova ricognizione ufficiale. La tomba viene scoperta il 6 dicembre e le ossa prelevate per un esame. A condurlo è il prof. Giovanni Canestrini, docente di anatomia comparata dell'Università di Padova. Nel corso di questa ricognizione, il cranio del Poeta si riduce in frantumi. (non era un gran problema perchè poteva essere restaurato)

1943: La guerra consiglia di mettere in salvo i resti del Poeta. L'apertura dell'arca per prelevarvi il contenuto avviene il 23 novembre e i resti vengono segretamente traslati a Venezia nei sotterranei di Palazzo Ducale, protetti da grandi lastre di marmo

1946: Finita la guerra, nei primissimi mesi del 1946, le ossa vengono ricomposte nella Scuola di Anatomia dell'Università di Padova. Il 26 giugno, i resti del Poeta vengono solennemente riportati nell'arca marmorea ad Arquà.

2003: In occasione del settimo centenario della nascita del Poeta, un gruppo di esperti viene incaricato dal Comune di Arquà e dalla Fondazione Cassa di Risparmio di Padova e Rovigo di una ricognizione scientifica finalizzata a verificare lo stato di conservazione dei resti (in relazione anche a possibili attacchi batterici) e a ricostruire il vero volto del Poeta (con tecniche particolari è possibile attraverso il cranio l'aspetto della faccia).

Al prof viene chiesto di recuperare il DNA di un frammento post craniale e un frammento craniale e confrontarli per vedere se appartiene allo stesso individuo.

Perché chi si era messo ad analizzare il cranio si insospettì sulla forma del cranio → perché potrebbe appartenere ad una donna.

Mentre c'erano rilievi patologici sullo scheletro (post-craniale) hanno dimostrato che effettivamente poteva essere attribuito al Petrarca: Frattura del Femore sinistro

Osseficazione di inserzioni ai tendini i cui sintomi sono ricollegabili alle descrizioni offerte da un carteggio tra il medico Giovanni Dondi ed il Petrarca.

Quindi il post craniale era probabilmente di Petrarca. Siccome lo scopo era quello di ricostruire il volto → volevano essere sicuri che il cranio apparteneva allo scheletro.

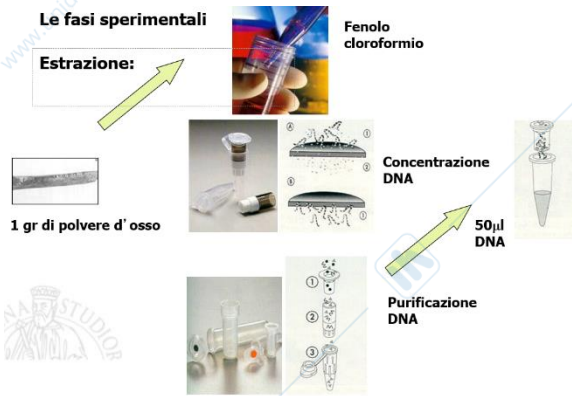
Se il DNA recuperato da frammenti ossei del post craniale e dal cranio avesse presentato la stessa sequenza di basi allora avremmo concluso che si trattava della "stessa persona" se invece le sequenze fossero state differenti probabilmente lo scheletro rinvenuto era una chimera!!!

Come si effettua uno studio sul DNA antico con la metodologia classica? (Passaggi viste lezioni fa)

Passaggi eseguiti: nella diapositiva c'è un dente, un frammento di costa e micropunta:

1. Si effettuano piccoli carotaggi dal frammento osseo e successivamente la "carota" viene polverizzata;

Il tutto viene eseguito in condizioni di massima sterilità.



Estrazione: con la metodologia classica si partiva da 1 g di polvere d'osso → Estrazione eseguita con Fenolo cloroformio → concentrazione del DNA → purificazione → 50 microlitri di DNA estratto.

La fase successiva all'estrazione del DNA è quella dell'amplificazione, attraverso una reazione biochimica che si chiama PCR → clonaggio → In questo modo si creano tante copie di DNA tutte uguali fra loro ;

Il DNA a questo punto è pronto per essere letto!!

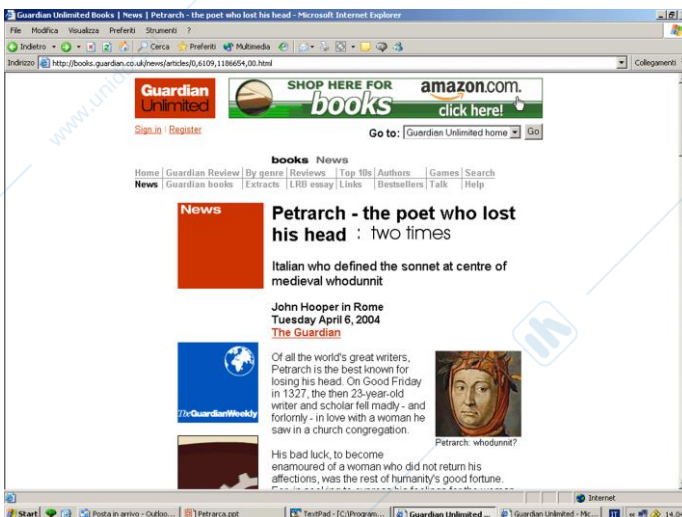
Letture che viene eseguita grazie ad un'altra reazione biochimica "reazione di sequenza" che si sviluppa in due fasi:

- una di laboratorio
- una davanti ad un Computer → sequenziamento che al computer viene ricostruita.

Risultati: i due frammenti (post craniali e craniali) hanno restituito due differenti profili di DNA mitocondriali → due individui differenti.

Slide 16ant5 → il profilo che sta sopra è differente da quello che sta in basso: nel primo profilo si osserva solo una mutazione puntiforme, il secondo ha 3 mutazioni puntiformi → QUINDI I DUE FRAMMENTI MITOCONDRIALI SONO DIFFERENTI → provengono da due individui tra loro differenti.

Giornale Guardian: Questo risultato ha suscitato molto interesse sia negli ambienti accademici che mass mediatici





Available online at www.sciencedirect.com

Forensic Science International xxx (2007) xxx-xxx



www.elsevier.com/locate/forsciint

WORK FLOW DI QUESTO ARTICOLO: è STATA UTILIZZATA LA METODOLOGIA CLASSICA. (vedi articolo).

Genetic analysis of the skeletal remains attributed to Francesco Petrarca

David Caramelli^{a,*}, Carles Lalueza-Fox^b, Cristian Capelli^c, Martina Lari^a,
 María Lourdes Sampietro^d, Elena Gigli^a, Lucio Milani^a, Elena Pilli^a, Silvia Guimaraes^a,
 Brunetto Chiarelli^a, Vito Terribile Wien Marin^e, Antonella Casoli^f, Roscoe Stanyon^a,
 Jaume Bertranpetit^b, Guido Barbujani^g

Materiali e metodi (vedi articolo) -FLORENCE

1. 'DNA extraction
2. Racemizzazione aa
3. UNG treatment: rimuovo le deaminazioni da parte della citosina
4. Quantificazione delle molecole di DNA → tramite RT-PCR
5. Amplificazione del DNA mit: le 360 bp di HVR1 è stata suddivisa in 3 frammenti utilizzando 3 coppie di primer.
6. Clonaggio e sequenziamento: è stato fatto con TOPO TA cloning kit(invitrogen)
7. Long'' amplified detection(Amplificare frammenti lunghi di DNA)e avere un risultato positivo è contro l'autenticità del campione degradato → perchè il DNA ha frammenti piccolo → quindi avere frammenti lunghi significa avere il campione contaminato.
8. Determinazione del sesso molecolare: verificare se il cranio appartenesse ad una donna. È stato fatto utilizzando il gene omologo dell'amilogenina e il frammento di 93 bp del cromosoma Y (SRY).

→Esperimenti ripetuti a Barcellona.

Risultati → si ottengono gli stessi risultati del frammento di HVR1 del DNA mit.

F1.1 F (Firenze) 1 (PCR n 1) 1 (estratto n 1.)

Se leggiamo F1.2 → estratto n1 e seconda PCR.

Se leggiamo F2.1 → estratto n.2 prima PCR

Perchè si fanno PCR multiple? Per confrontare il risultato e confermare i risultati

Stesso per B.1.1 ecc.. -> confermare i risultati nel 2° laboratorio → sono state ottenute nello stesso punto la stessa mutazione puntiforme.

Le sostituzioni sparse sono misincorporazioni.

DNA mit ci ha detto che erano due individui differenti → è stato determinato il sesso → post craniale: maschio. Craniale: femmina

Non si sono accontentati → hanno fatto anche un'altra cosa → datazione radiometrica al carbonio 14 della costa e del dente e si sono accorti che il dente era più vecchio di 200 anni rispetto alla costa → quando è stato riassembleto lo è stato fatto con un cranio completamente diverso.

CRS (cambridge ...) → sequenza di riferimento universale depositata in genebank.

RCRS → sequenza di riferimento che è stata assemblata → sequenza di riferimento. LA differenza è che la seconda è un mitocondrio completo.

Secondo studio: **PALEOGENETIC AND MORPHOMETRIC STUDIES OF MAN**

È stato dimostrato che l'uomo di Altamura era un Neandertal. Prima gli studi erano solo di carattere morfologico → non ha permesso l'identificazione come Neandertal.

Campionamento in data 16 luglio 2009: si trova in una cava (grotta) in Puglia → scheletro ricoperto di carbonato di calcio perchè è una grotta molto umida → DNA molto frammentato perchè la grotta è umida.

Per il campionamento è stato utilizzato una pinza (come quelle delle sale giochi) → Sono state prese delle ossa che stavano dietro il cranio dell'uomo di Altamura. C'erano dei fori e grazie ad una sonda e una specie di pinzetta è stato recuperato frammenti ossei (frammenti di scapola) e messa in un sacchetto sterile → laboratorio.

In laboratorio:

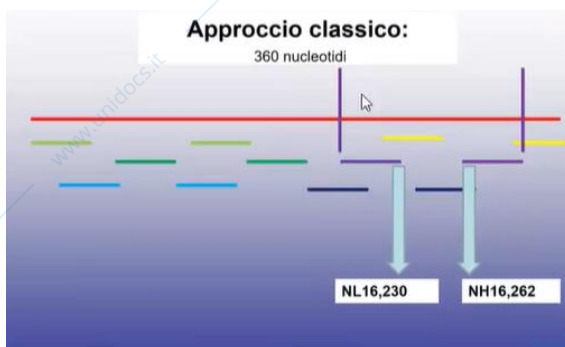
Taglio e pulizia del reperto (2009)

Amplificazione della reg HVR1 sappiamo che i neandertal e sapiens per questa regione hanno delle differenze che possono essere ben osservabili. Nell'HVR1 di un Neandertaliano osserviamo dalle 21 alle 27 sostituzioni nucleotidiche dal Sapiens.

Due differenti approcci

- 1) classico
- 2) innovativo, NGS

APPROCCIO CLASSICO: più coppie di primer neandertaliani → amplificazione dei 360 nt perchè il DNA è molto frammentato → sono stati utilizzati molte coppie di primer perchè c'erano molti frammenti piccolo perchè la grotta era umida. Infatti è stato amplificato un piccolissimo frammento con coppia primer → 16262-16260 → da un punto di vista diagnostico ci permetteva di ascrivere quell frammento di DNA mit a Neandertal.



se queste mutazioni le andiamo a confrontare con altri neandertaliani → ha le stesse mutazioni diagnostiche di tutti gli altri neandertaliani → abbiamo potuto dire che si trattava di un neandertal.

È stato ricostruito un albero filogenetico → il neandertal di alta mura cade nella variabilità genetica dei neandertaliani italiani e spagnoli per questo frammento diagnostico.

LEZIONE 28 OTTOBRE

NGS ANALISI DEI RESTI UMANI ANTICHI TRAMITE il flusso lavoro NGS.

In tutti i laboratori in cui si fanno studi di carattere antropologico, in cui le molecole come il DNA sono la base fondamentale su cui si sviluppano le ricerche → avere un approccio NGS significa strutturare un'attività di ricerca che si complementa tra antropologi e fisici, tra antropologi molecolari e tra ricercatori che si occupa di bioinformatica → significa avere un aspetto multidisciplinare in questo orizzonte di ricerca: è un orizzonte di ricerca all'avanguardia. queste ricerche possono essere declinate in aspetti di carattere forense e biosanitario.

FLUSSO DI LAVORO NGS

1. La prima fase è l'**estrazione del DNA** → si sta parlando di analizzare genomi completi, si parte da reperti ossei che non sono post-craniali, ma stanno in distretti ossei ben specifici:

- rocca petrosa (osso sotto l'orecchio) e denti. Estrarre il DNA da rocca petrosa non è semplice → Rocche petrose sono attaccate al cranio → quindi nell'estrarle possono essere rotte. Anche se ci sono tecniche che sono meno invasive, ad oggi le rocche petrose divisa in due → Il DNA viene recuperato dal punto C. tutto ciò in condizioni di massima sterilità.

- Denti → buco nella radice → resa minore ma più semplice della rocca petrosa.

Recupero del campione: con NGS si devono recuperare questi tipi di campioni, perché solo attraverso questo si hanno genomi completi.

Con la metodologia classica noi ci possiamo focalizzare su alcune tipologie di individui, con l'NGS possiamo andare a studiare tutti i genomi che vogliamo andare a studiare, perché grazie alla tecnologia NGS → si può distinguere DNA endogeno da esogeno → si può estrarre qualsiasi tipo di campione manipolato da qualsiasi tipo di professionista → e grazie al flusso di lavoro a cui si affida la tecnologia NGS si può distinguere il DNA esogeno da endogeno e quindi scartare campioni molto contaminati. Grande passo avanti!

Attraverso la tecnologia NGS è possibile applicare anche a caratteri di tipi forense: discriminare un mitocondrio, per vedere se è contaminato o no (se appartiene a più fonti biologiche).

Prima di parlare di NGS

Cycle sequencing: è un metodo di Sanger modificato → metodologia di sequenziamento del DNA basato su PCR modificata, abbiamo reagenti simili ma con qualche modifica.

- dNTP
- primer
- Taq polimerasi
- DNA
- ddNTP (dideoossi): dideossinucleotidi o ddNTP, sono deossiribonucleotidi sintetici che presentano lo zucchero desossiribosio privo del gruppo ossidrilico in posizione 3'. Il deossiribosio manca già del gruppo ossidrilico in posizione 2'p, ma hanno un fluoroforo specifico per ogni base. → fluoroforo emette un determinato colore quando colpita da un determinata lunghezza d'onda. Il fatto di non avere un gruppo OH comporta che ogni volta che viene inserito dalla Taq pol nella polimerizzazione del DNA da sequenziare → il ddNTP non riesce ad attaccare né un ddNTP e né un dNTP → la catena si ferma.

Il DNA da cui si parte nella reazione di sequenza deriva dalla PCR (in tante copie), se no non sarebbe possibile sequenziare il DNA che voglio andare a leggere. (vedi tesi)

Denaturazione del DNA, annealing del primer e estensione della catena. Dove viene inserito un ddNTP la catena si interrompe e non va avanti con formazione di molti frammenti. Il tutto è caricato in un gel elettroforetico (anche se oggi esistono sequenziatori automatici che hanno capillari), i frammenti corrono verso il polo positivo e secondo la loro grandezza.

Avviene elettroforesi → si ottiene un elettroferogramma che ci permette di ricostruire la catena della sequenza del DNA.

(È il sequenziamento di sanger fatto con cycle sequencing, cioè non con 4 tubicini differenti ma con un unico tubicino in cui metto tutti e 4 i dNTP, è una modifica)

PROSPETTIVA MITOCONDRIALE E NUCLEARE CON METODOLOGIE DI NUOVA GENERAZIONE

Perché ci sono queste due prospettive? Perché con il flusso NGS si va a capire come si vanno a ricostruire i mitocondri e i genomi mitocondriali.

È importante la prospettiva mitocondriale → L'analisi dei mitocondri (ci permette di fare la storia evolutiva a livello femminile, aplotipi e aplogruppi ecc.) e l'analisi della conservazione dei mitocondri ci permette di capire che probabilità abbiamo di ottenere un buon DNA nucleare.

NGS (NEW GENERATION SEQUENCING) revolutions.

Due motivi per cui si chiama NGS revolution → perché con NGS anziché leggere piccoli frammenti, si leggono interi genomi → è una rivoluzione nel campo dell'antropologia e non solo → panoramica ampia per le analisi molecolari

Con NGS si va ad analizzare la rocca petrosa.

Ci possono essere una serie di passaggi che ci danno un'affidabilità maggiore e più semplici da fare.

Il problema nell'analisi NGS è che solo alla fine del percorso, con la lettura delle sequenze si capisce se si deve scartare o no il campione. È vero che possono essere fatte analisi preliminari, ma queste sono analisi che prevedono il sequenziamento NGS. Alcune analisi preliminari ci danno anche info importanti. Es. determinazione del sesso.

Con Next generation sequencing si passa dal 2004: abi sanger sequencing (10^2 frammenti per corsa) fino al NovaSEQ, ultimo strumento di sequenziamento prodotto dalla casa produttrice ILLUMINA → riesce a sequenziare 6 tera (6 mila miliardi) di genoma.

Tutto questo è iniziato nel 2005 con 454 GS20 detto anche pirosequenziamento → ci ha permesso di fare i primi studi sul genoma dell'homo di Neandertal (1 milione di paia di basi)

SVILUPPO DEL Flusso di lavoro NGS con la 454 utilizzata anche in ambito forense, anche se ora usano ILLUMINA.

1) Estrazione del DNA

2) Costruzione libreria a DNA → immortalizzare *tutto* il DNA che è contenuto all'interno di un estratto, quindi su tutto ciò che andiamo ad estrarre (DNA batterico, endogeno, esogeno ecc.).

Il 454 flx è stato utilizzato per analizzare un milione di bp nell'Homo di Neandertal, ma quando la sua tecnologia da un punto di vista bioinformatico è stata ottimizzata ci ha permesso di ricostruire mitocondri

completi. Il primo sequenziamento sui mitocondri completi dei neandertaliani è stato fatto con la 454 → è NGS di prima generazione, che permette di analizzare grandi quantità di DNA

Costruire una libreria → significa avere tanti frammenti di DNA di differenti dimensioni a cui si aggiungono adattatori che sono oligonucleotidi (piccoli frammenti di DNA noti): adattatore A e adattatore B (DNA precedentemente denaturato)

WORK FLOW 454: faccio interagire la libreria con delle microbiglie che hanno la caratteristica di avere sulla loro superficie degli adattatori A o degli adattatori B complementari.

È importante stimare la concentrazione della libreria (con PCR-RT) per capire il numero di molecole che ho generato all'interno della mia libreria: è importante perché devo far reagire la mia microsfera con un frammento della mia libreria → Se nella libreria ho 1 milione di molecole devo utilizzare 1 milione e 100 mila di sferette per avere la fortuna e consapevolezza che casualmente ogni sfera reagisce con una molecola → quindi utilizzare un numero > di biglie rispetto alla libreria.

Una volta che si attacca il frammento della libreria alla sfera → si ha la polimerizzazione del frammento di libreria → **PCR emulsionale su ogni sferetta** che permette la riproduzione del frammento inserito: . La ricostruzione di questo frammento che avviene su più adattatori attaccati alla sfera → si attacca → polimerizzazione → si stacca → si riattacca su un altro adattatore. La PCR dura fin quando ci sono gli adattatori sulle sfere → Una volta che è avvenuta l'amplificazione della PCR emulsionale → si rompe emulsione → queste microsferette vengono inserite in piastre dove ci sono forellini (ca milioni su queste piastre) che si chiamano microreattori. In ciascun forellino vengono messi reagenti per il pirosequenziamento (utilizzata solo per questa tecnologia).

→ In ciascun microreattore ho la lettura di un frammento della libreria → quindi ho il panorama molecolare della libreria che ho estratto.

DOMANDA. Secondo noi questo sistema in che passaggio lo posso inserire nella metodologia classica? O meglio il pirosequenziamento cosa potrebbe sostituire nel work flow della metodologia classica? il clonaggio → il pirosequenziamento come tecnologia NGS ti permette di avere un panorama molecolare, cioè leggere tutte le molecole che sono presenti in un determinato contesto (dall'amplificazione o dall'estrazione).

Posso andare a sequenziare con il pirosequenziamento il prodotto di amplificazione → vedo tutte le mie molecole che composizione hanno.

Si può utilizzare anche in ambito forense? E su cosa posso utilizzarlo? Abbiamo parlato di DNA fingerprinting che si fa utilizzando gli STR (polimorfismi di lunghezza) → due gemelli monozigoti hanno lo stesso DNA fingerprinting in termini di lunghezza ma nel corso della loro vita potrebbero avere sviluppato delle mutazioni sugli STR che se vado a fare il polimorfismo di lunghezza non osservo.

Ma se amplifico gli STR e sequenzio il prodotto di amplificazione → vedo la lunghezza degli STR e la sequenza. Quindi il pirosequenziamento può essere utilizzato per andare ad osservare i polimorfismi di sequenza e di lunghezza.

Quindi con la tecnologia 454 posso sequenziare DNA che provengono da estratti, da ampliconi andando a costruire librerie e leggerne il panorama molecolare → prima tecnologia NGS.

1) preparazione libreria (anche ampliconi) → frammentazione DNA. I due adattatori A e B si legano a ciascuna estremità → dopo il filamento si denatura → un filamento si lega ad ogni biglia

Se ho una biglia con due filamenti → scarto la biglia.

2) preparo PCR emulsio-clonale → biglia inserita nell'emulsione dove all'interno avviene la PCR → il filamento si riproduce finché la biglia è saturata → viene rotta emulsione → ciascuna biglia che ha tutti i filamenti riprodotti uguali a se stessi

→ caricati sulla piastra dove ogni pozzetto c'è solo una biglia → inseriti i reagenti per il sequenziamento → caricati in macchinario 454 → LCD camera che legge la fluorescenza emessa da ciascun frammento → programmi che riportano la sequenza del determinato frammento che ho sequenziato → **si ottiene il panorama molecolare di una libreria.**

LEZIONE 02 NOVEMBRE

Ora vediamo la seconda applicazione della tecnologia NGS relativa alla FLX per ricostruire i mitocondri nell'HOMO di Neandertal.

Lavoro fatto nel 2009, in una pubblicazione su Science andava a ricostruire 5 mitocondri completi di 5 differenti neandertaliani utilizzando come step finale il sequenziamento con la 454 FLX.

Possono essere sequenziati non solo prodotti di amplificazione, ma anche frammenti di DNA recuperato nel contesto dell'estrazione.

Tecnologia PEC (primer extension capture) che ci permette di andare a selezionare il DNA che vogliamo successivamente sequenziare. È stata utilizzata per la prima volta in questo lavoro.

Come viene ricostruita una libreria: si inserisce su un DNA che ho estratto, amplificato ecc.. i due adattatori (A e B) che ci permettono di sviluppare il sequenziamento su questi beads (biglie) sulle quali sono inseriti gli adattatori complementari agli adattatori presenti sui frammenti di DNA per costruire la libreria.

Si possono taggare le librerie → significa che oltre agli adattatori possono essere inseriti per una particolare libreria un frammento di DNA specifico noto, insieme all'adattatore → successivamente quando andrò a leggere il sequenziamento della libreria saprò che appartiene a quel determinato campione.

Se io taggo una libreria significa che posso mescolare assieme più librerie con diversi tag (con diverse sequenze che contraddistinguono quella determinata libreria) e posso sequenziarle tutte insieme → oltre a leggere la sequenza leggo anche il tag → quindi so che appartiene a quella determinata libreria. Questo comporta che posso andare a sequenziare un numero maggiore di campioni tutti assieme e posso leggere i risultati del mio sequenziamento e attribuire ciascuna sequenza ad un determinato estratto e quindi ad un determinato campione.

Esistono diversi tipi di #TAG:

Nel work flow della 454 FLX: questi tag si chiamano MID → di questi MID ce ne sono più o meno una ventina → teoricamente posso processare 20 librerie assieme e andarle poi a riconoscere.

Questo è quello che è stato fatto in questo lavoro → costruzione di librerie che contenevano un TAG → con la 454 FLX riesco ad ottenere $10^{5,6,7}$ molecole sequenziate per corsa.

Supponiamo 10^6 (un milione), il DNA mit è 16000 (10^4) → Se voglio sequenziare un mitocondrio umano → un mitocondrio da solo lo leggo 100 volte → leggo ciascuna base 100 volte. ($10^6/10^4 =$

100 volte → leggo 100 sequenze → posso andare a leggere su ciascuna sequenza la base che trovo 100 volte.

Se un mitocondrio lo leggo 100 volte, due ne leggo 50 volte. Utilizzando il tag posso leggerli insieme.

Questo sistema è stato utilizzato in questo lavoro da Alan Brix(?) per analizzare e sequenziare 5 DNA mitocondriali insieme → la profondità di lettura è 20.

Profondità di lettura → quante volte leggo quel determinato frammento di DNA. Se lo leggo 5 volte → 5x ecc...

Il passo successivo è leggere DNA mitocondriale → utilizzo un sistema che si chiama **PEC** (primer extension capture): si basa su un principio che è di andare a catturare la libreria di DNA che ho estratto da un reperto neandertaliano → utilizzo PEC che mi permette di rimuovere dal contesto solo i frammenti che voglio sequenziare.

La PEC sarà sviluppata con piccoli frammenti di oligonucleotidi che vanno a catturare il DNA mitocondriale umano → Questi primer PEC sono frammenti complementari al frammento di DNA che voglio catturare.

COM'E' FATTO UN PRIMER PEC



Sequenza di DNA che è complementare ad un frammento di DNA mitocondriale, con coda (Ty) e infondo la coda c'è la biotina. Per catturare tutto il mitocondrio avrò la necessità di avere numerosi primer PEC che coprono tutta la sequenza del DNA mitocondriale.

Il DNA mitocondriale in un Neandertal è tutto frammentato di varie lunghezze.

FLUSSO DI LAVORO per la cattura di un mitocondrio neandertaliano.

Questo ci permette fare delle inferenze sulla struttura di popolazione dei neandertaliani.

ESPERIMENTO SULLA CATTURA DI UN MITOCONDRIO DI UN NEANDERTAL

Si parte da frammento di Rocca petrosa (2) e denti (2) di un Neandertaliano → su questi 4 frammenti ossei dopo averli sterilizzati puliti ecc.. → si recupera polvere d'osso → estrazione del DNA (provette con DNA estratto) → costruzione libreria dove ognuna deve contenere i due adattatori e un tag → ho quindi una libreria su tutto il DNA estratto → utilizzo primer PEC per catturare DNA mitocondriale → si attacca alla zona complementare dove alla fine del primer c'è la biotina → si estende a tutta la catena → ho un complesso PRIMER PEC - frammento che ho catturato → rimuovo il frammento che ho catturato utilizzando delle microsferiche magnetiche che hanno sulla loro superficie la streptavina che si complementa alla biotina → formazione complesso biotina-streptavina → utilizzo un magnete e vado meccanicamente a tirare via dalla mia libreria le molecole che a me interessano → dopo il recupero delle molecole che mi interessano → separo il primer PEC dalla molecola che mi interessa → le molecole ottenute seguono il flusso di lavoro della 454 FLX (del pirosequenziamento). → quindi andranno ad interagire con le biglie

magnetiche su cui ci sono gli adattatori complementari A e B → PCR emulsioclonale → biglie magnetiche nella piastra dove ci sono i microreattori → piastra nella macchina 454 → sequenziamento DNA mit.

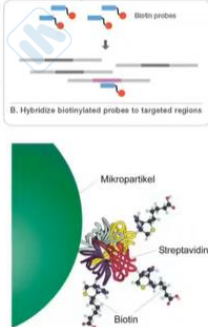
UNIVERSITÀ DEGLI STUDI FIRENZE

NUOVO APPROCCIO METODOLOGICO:

Target selection → Genome enrichment

per 'pescare' o '**catturare**' solo alcune specifiche regioni genomiche all'interno di una libreria NGS

- utilizza **sonde a DNA** complementari alle regioni genomiche di interesse per selezionarle
- sfrutta l'interazione tra le **molecole di biotina e di streptavidina** coniugate a supporti magnetici (biglie) per catturare solo i frammenti di DNA selezionati dalle sonde



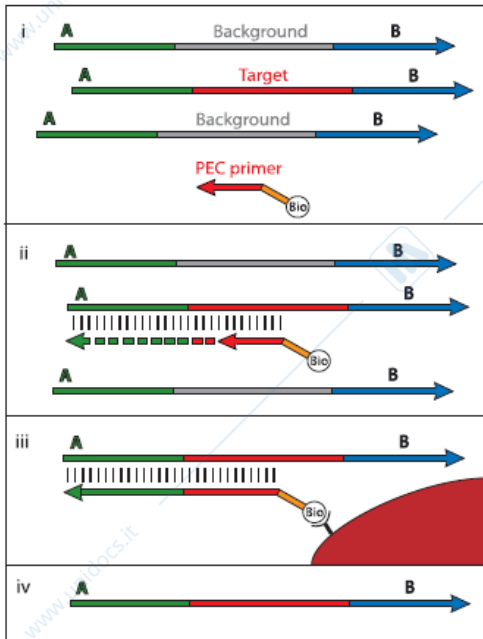
Molto spesso viene arricchita la libreria → cioè aumentare il numero di copie → viene amplificato, ma non con la PCR (perché lo fa in maniera esponenziale). → infatti molto spesso si parla di Target enrichment capture.

Vantaggi rispetto alla PCR:

- numero maggiore di target
- permette di analizzare anche frammenti molto corti (<50bp, fino a 30-20bp)
- Mantiene inalterate le caratteristiche molecolari delle molecole originali estratte dai campioni (lunghezza, siti di frammentazione)

PEC → Primer Extension Capture

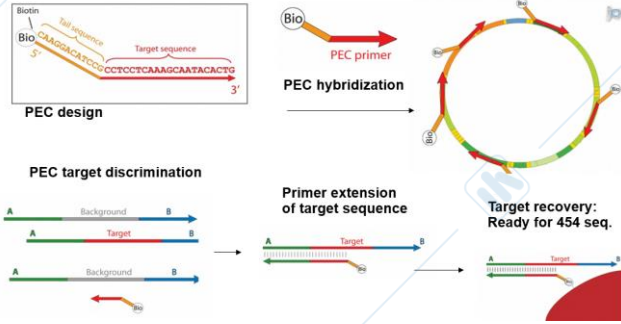
- Applicata al sequenziamento di 5 genomi mitocondriali neandertaliani
- 574 PEC primers per l'intero mtDNA



Primer Extension Capture (PEC)

Complete mitochondrial DNA (mtDNA) genomes of five Neandertals
 J Vis Exp. 2009 Sep 3;(31):1573. doi: 10.3791/1573.
 Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Paabo S.

<http://www.jove.com/video/1573/primer-extension-capture-targeted-sequence-retrieval-from-heavily?ID=1573>



<http://www.jove.com/video/1573/primer-extension-capture-targeted-sequence-retrieval-from-heavily?ID=1573>

Funzionamento: il background significa ciò che c'è nella libreria con gli adattatori, il target e il primer pec → il primer pec si complementa al target → c'è l'estensione → successivamente la biotina si complementa con la streptavina → cattura.

Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes

Adrian W. Briggs,^{1*} Jeffrey M. Good,¹ Richard E. Green,¹ Johannes Krause,¹ Tomislav Maricic,¹ Udo Stenzel,¹ Carles Lalueza-Fox,² Pavao Rudan,³ Dejana Brajković,⁴ Željko Kučan,³ Ivan Gušić,³ Ralf Schmitz,^{5,6} Vladimir B. Doronichev,⁷ Liubov V. Golovanova,⁷ Marco de la Rasilla,⁸ Javier Fortea,⁸ Antonio Rosas,⁹ Svante Pääbo¹

Analysis of Neandertal DNA holds great potential for investigating the population history of this group of hominins, but progress has been limited due to the rarity of samples and damaged state of the DNA. We present a method of targeted ancient DNA sequence retrieval that greatly reduces sample destruction and sequencing demands and use this method to reconstruct the complete mitochondrial DNA (mtDNA) genomes of five Neandertals from across their geographic range. We find that mtDNA genetic diversity in Neandertals that lived 38,000 to 70,000 years ago was approximately one-third of that in contemporary modern humans. Together with analyses of mtDNA protein evolution, these data suggest that the long-term effective population size of Neandertals was smaller than that of modern humans and extant great apes.

Articolo di questo lavoro (analisi sui 5 mitocondri neandertaliani).

Risultati dell'analisi (leggi).

Video da sapere all'esame: media di cattura 55x → riguarda.

Quindi la PEC è uno strumento che permette di catturare genomi, nel caso specifico erano genomi neandertaliani e attraverso ciò hanno ricostruito il DNA mitocondriale, ma anche le dinamiche di popolazione, la grandezza della popolazione dei neandertaliani e la loro variabilità.

LEZIONE 9 NOVEMBRE

Articolo: a complete mtDNA genome of an early modern human from Kostenki, Russia

È stato fatto il DNA mit di un individuo antico umano sapiens che viveva in Asia, 40 mila anni fa. È stato osservato che il DNA mit di questo individuo cadeva nella variabilità genetica delle popolazioni attuali → pertanto era un'ulteriore riprova che i primi sapiens che hanno lasciato l'Africa e sono arrivati in Europa non condividevano la stessa sequenza mitocondriale neandertaliana. ??

Quello che interessa è la problematica → Con metodologia classica, recuperare genomi umani di individui vissuti nel passato era difficile soprattutto non conoscendo la storia tafonomica → subentra una problematica che impediva di poter andare ad analizzare tanti individui rinvenuti da scavi archeologici che ci permettono di capire di più le dinamiche della storia evolutiva.

In questo lavoro si dimostra per la prima volta come sia possibile analizzare genomi di individui umani vissuti nel passato e di verificare la bontà del dato → cioè la quantità di DNA endogeno presente nel nostro campione e di scartare o tenere quel risultato in base alla quantità di DNA endogeno del reperto che vado ad analizzare.

Ora vediamo come capire se siamo di fronte a DNA esogeno o endogeno.

1) **COSTRUZIONE LIBRERIA**: recuperando il DNA di un individuo antico, siccome il DNA è frammentato → non avrà le estremità piatte, per poter inserire gli adattatori → devo riparare le estremità delle molecole tramite T4 polimerasi a livello del 5' o 3' → ottengo estremità piatte.

Sappiamo che il DNA è degradato può avere misincorporazioni → T4 quando inserisce un nucleotide, se il nucleotide stampo è deaminato, T4 non lo riconosce e commette un errore. Es. Deaminazione citosina → uracile. Quando T4 passa mette una timina. **È stato visto che in un DNA degradato le misincorporazioni in 5' e 3' molto frequenti → in particolare in C → T** sono molto frequenti proprio perché devo ricostruire la parte mancante.

L'aspetto interessante era: **Come associare questa frequenza di C → T nelle posizioni 5' e 3' ad un campione degradato?**

Andando a osservare le caratteristiche dei mitocondri neandertaliani → Nei mitocondri neandertaliani le C → T erano molto elevate, fino a 50% di queste misincorporazioni. **Ma per essere sicuri che queste misincorporazioni erano caratteristiche di un DNA antico** → avevano analizzato il mitocondrio dei neandertaliani e siccome in termini di sequenze riconosciamo le differenze tra un mit umano e uno neandertaliano → siccome queste differenze erano presenti sui mitocondri neandertaliani → **queste C → T erano un indicatore di DNA antico.**

(hanno visto che i neandertaliani hanno in 5' → 3' le misincorporazioni C → T → quindi hanno pensato che fosse un segnale di antichità del DNA. Perché proprio in pos 5' 3'? → perché è degradato → è frammentato → perché proviene da campione antico → ha la capacità di subire misincorporazioni in quei punti dovuto all'azione dell'ossigeno, umidità ecc...)

Per ricostruire un frammento di DNA mit → devo avere molte letture per un determinato frammento → per avere molte letture devo avere la cattura con la PEC → sequenziamento della singola molecola che ho catturato.

Un campione antico ha ca 20-30% di misincorporazione in 5' e 3'. Quindi significa che su 100, 20 sono misincorporazioni C→T, gli altri 80 no.

Come faccio a sapere se il resto sono antichi o contaminanti? Le distinguo da contaminanti dal motivo della sequenza: possono esserci molecole che non hanno C→T ma hanno lo stesso motivo della sequenza con C→T.

Quindi si distingue il dna esogeno da endogeno:

- C→T
- i pattern di frammentazione → DNA molto frammentato è un DNA antico. Anche se quando si vanno a costruire le librerie anche un DNA lungo viene frammentato.

Tutto questo lavoro l'hanno potuto fare perché hanno potuto catturare il DNA neandertaliano con la PEC.

Inoltre le deaminazioni e quindi le misincorporazioni AUMENTANO con il passare del tempo → relazione pattern di misincorporazione in 5' 3' a seconda dell'antichità del campione → tanto più il campione è vecchio → tanto più è elevato il pattern di misincorporazione.

Questo ci permette di essere sicuri di analizzare del DNA endogeno su campioni degradato → quindi di recuperare informazioni che ci permettono di fare delle inferenze sui genomi degli individui vissuti nel passato e quindi di ricostruire la storia evolutiva.

Per fare tutto ciò si utilizzano software: **MAP damage** che legge le misincorporazioni in 5' 3' su un pattern di reads che sono generate da una libreria che viene sequenziata con NGS.

Ritorno sull'articolo

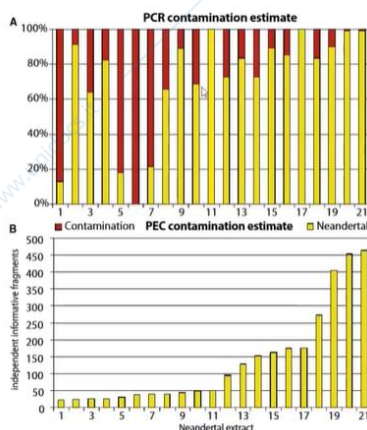
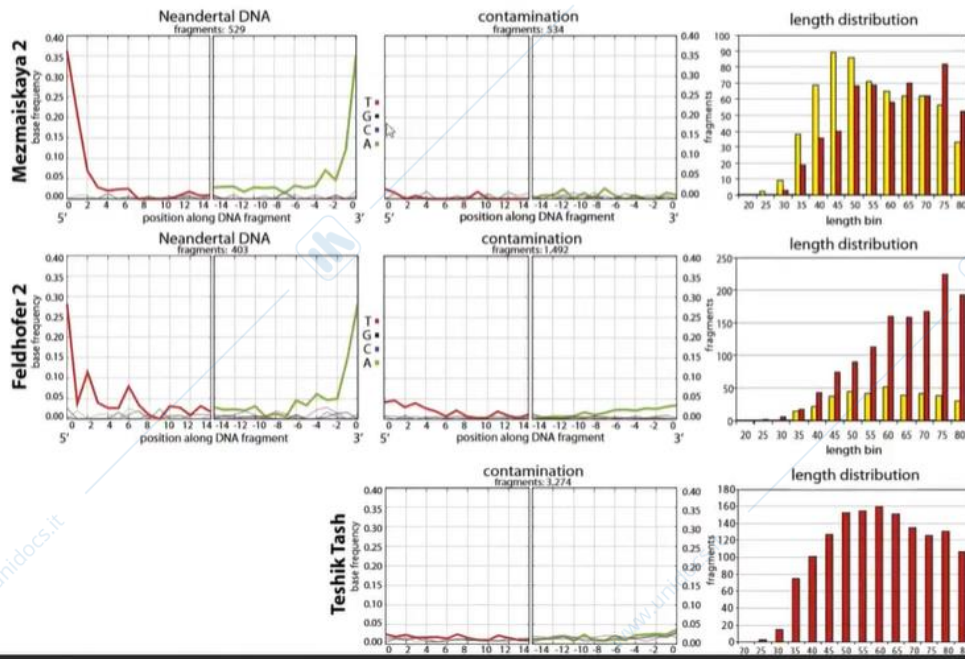


Grafico: Rosso: Dna contaminante, giallo: DNA neandertaliano.

Come abbiamo fatto a comprendere quale era il DNA contaminante e quale neandertaliano? Con la PEC abbiamo catturato il DNA neandertaliano e contaminate. Quest'ultimo è stato escluso perché il motivo del DN era una sequenza non neandertaliana.



1 campione (Mezmaiskaya2) è Neandertal → C→T in 5' e A → G in 3' intorno a 35% nei campioni gialli che sono quelli neandertaliani.

La frequenza delle C→T nelle reads non neandertaliani → vicini allo zero.

Viene messa anche la distribuzione della lunghezza → % della lunghezza delle READ 45-50 bp dei neandertaliani

% Lunghezza del DNA contaminante è maggiore, ca 75 bp.

2 campione (Feldhofer): altro neandertal → stessa cosa.

3 campione: teshik tash, reperto neandertaliano trovato in Iran → ha estremità piatte, quindi in 5' 3' non ci sono punte. Guardando la lunghezza → la maggior parte di tratta di sequenze contaminanti → sapiens. Le riconosco dal motivo mitocondriale neandertaliano che è differente da quello di sapiens.

Articoli dove hanno stimato i pattern di degradazione analizzando i C→T analizzando campioni di differenti età.



tanto più un campione è antico → maggiore sono le misincorporazioni → campioni più recenti (da 15 a 20%).

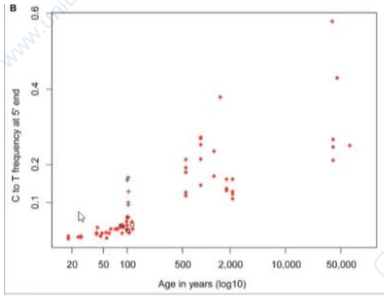
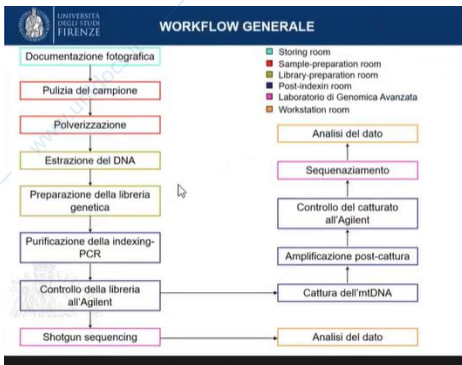


Figure 4. Characteristics of C to T misincorporations. A: C to T misincorporations at the first 15 bases of endogenous mtDNA fragments from a 500-600 year old horse sample (Sample 54). B: C to T misincorporations at the first positions of mtDNA fragments as a function of age. New samples known to have been "roasted" over fire and treated with ponal glue are indicated by crosses and four samples treated by the "Leipzig cocktail" are indicated by circles. doi:10.1371/journal.pone.0034131.g004

Si può notare che tanto più un campione antico → tanto più alte sono le misincorporazioni.

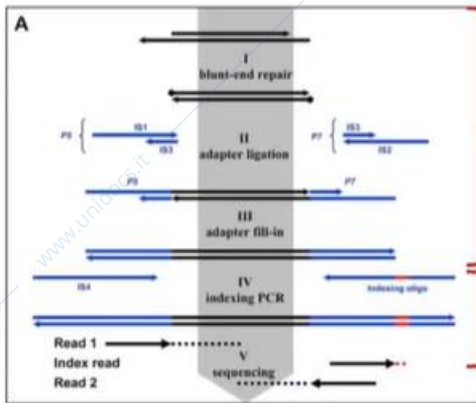
Le C→T ci indicano oltre alla degradazione del campione, ma in un arco temporale abbastanza grande si sa che tanto più sono antichi → maggiore è il livello di misincorporazioni.

METODICHE PER LO STUDIO DEL DNA ANTICO



COME SI COSTRUISCE UNA LIBRERIA DI ILLUMINA:

protocollo si trova nell'articolo: illumina sequencing library preparation for high multiplex target capture and sequencing.



Nella libreria illumina, i TAG si chiamano **INDICI** o **INDEX**.

1) parto da un frammento di DNA con estremità sfalsate → il primo passaggio è **blunt and repeat**: ricostruisco le estremità mancanti con formazione di estremità piatte → quindi inserisco gli errori

2) **adapter ligation** → inserisco adattatori che sono di differente lunghezza → p5 e p7 → ligasi che attacca i due adattatori.

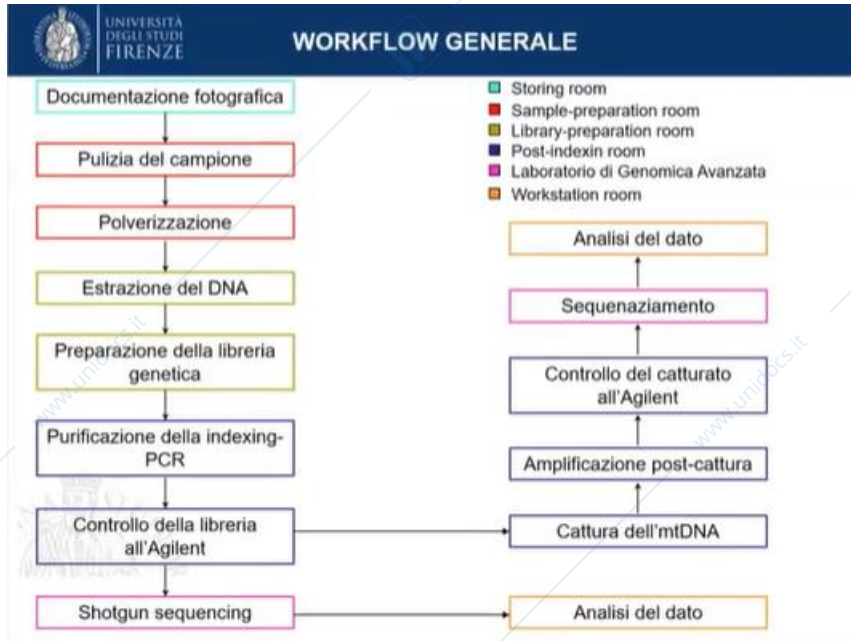
3) **Adapter fill-in** → si estendono le porzioni degli adattatori ancora a singolo filamento, la reazione è operata dalla Bst polimerasi → ottengo estremità piatte

4) **Indexing PCR** → con una PCR con pochi cicli dove si inseriscono gli indici che permettono di fare highly multiplex sequencing → sequenzio più campioni insieme perché li riconosco attraverso gli indici.

5) Sequenziamento

LEZIONE 11 NOVEMBRE

WORK FLOW GENERALE per analizzare un individuo vissuto nel passato con tecnologia di nuova generazione.



1) Documentazione

fotografica: per documentare la tipologia del campione, e l'etichetta (label) del campione. questa fase viene fatta nella *storing room*

Molti degli errori che vengono fatti riguarda lo sbaglio nel mettere una giusta etichetta al nome del campione.

Es: Recupero un campione proveniente da uno scavo archeologico: Ripatetta 1 e Ripatetta 9.

Ripatetta è uno scavo che si

trova in puglia dove ci sono diversi campioni numerati. Su questi campioni viene aggiunto anche lo strato in cui vengono trovati → si ha il nome del campione costituito dal: nome del luogo, sigle che l'archeologo da a questo campione). → non possiamo processare un campione con un nome complesso/lungo.

In laboratorio ad ogni reperto viene data un ID che è in ordine progressivo: ad esempio al campione Ripatetta 1 viene dato il n. 1 + una sigla dove è stato recuperato il reperto. → quindi deve esserci un archivio dei reperti che entrano in laboratorio in cui vengono siglati tutti i campioni, e quindi un ID che viene utilizzata in lab. per descrivere il campione. Questa identità deve riferirsi al nome del campione fornito dall'archeologo. **Buona documentazione fotografica** che descrive il nome del reperto con cui arriva in laboratorio e le varie fasi successive del ricercatore che mette le sigle sul campione che entra nel flusso di lavoro delle analisi di laboratorio.

2) **Pulizia del campione:** si rimuove la parte superficiale e polverizzazione nella stanza "*sample preparation room*" → luoghi nei quali non sia avvenuto in contatto in grandi concentrazioni.

3) **estrazione del DNA:** in soluzione la molecola di DNA con delle particolari matrici che sono ottimizzati per recuperare frammenti di DNA molto corti. (perché se il DNA è altamente degradato)

e preparazione della libreria genica nella stanza "*library preparation room*" → è una stanza che non è stata sottoposta a processi di amplificazione di DNA.

4) **Purificazione della indexing-library:** viene fatta la purificazione della librerie e la indexing-PCR

5) **Controllo della libreria all'Agilent:** agilent è un elettroforesi molto specifica che va ad indicare la lunghezza dei frammenti della libreria e quindi di quantificare la lunghezza dei frammenti: sappiamo che

partiamo da DNA degradato e aggiungiamo dei frammenti (adattatori) di lunghezza nota. Se estraggo dei frammenti da un campione altamente degradato → avrò frammenti di 50-60 bp → se aggiungo adattatori di 20 bp e indici di 10 bp → avrò una media di frammenti di lunghezza di ca 100-120 bp. Questa fase avviene in indexing room

A questo punto posso andare a sequenziare tutto il genoma che ho in una libreria, oppure posso andare a catturare il DNA mitocondriale (tramite PEC).

Ci sono altre metodiche per andare a catturare regioni specifiche più economiche e efficienti: **PROTOCOLLO MARECIC**: per la cattura di DNA mitocondriali e di genomi nucleari.

Quali sono i frammenti ossei da cui si parte per fare queste analisi?

La rocca petrosa è il frammento osseo da cui si riesce ad ottenere la maggior parte di DNA anche in buono stato di DNA.

I denti sono frammenti ossei da cui si ottiene una discreta quantità di DNA.

Ossa lunghe, ossicini dell'udito, tartaro, semi antichi, coproliti (feci fossili → ci permette di associarli anche ad un particolare tipo di dieta, quindi stile di vita di una particolare popolazione).

a) Osso spugnoso
b) Zona un po' più densa attorno alla capsula otica
c) Capsula otica, zona estremamente densa e compatta

Il frammento osseo viene recuperato dal punto c (capsula otica).

Come la campioniamo?

- 1) Distacco della rocca petrosa dal cranio
- 2) Taglio della rocca e polverizzazione con microtrapano nella zona C, dove l'osso è più compatto.

Non può essere fatto su tutti i crani, perché ci sono crani molto importanti. Ci sono dei sistemi che si basano su una tomografia assiale computerizzata (una TAC al cranio) → viene individuato il punto C dall'esterno e con un microtrapano viene prelevata la polvere d'osso. In ambito forense è molto importante perché in un reperto scheletrico da cui si vogliono recuperare informazioni dal punto di vista genomico si deve recuperare DNA dalla rocca petrosa.

3) Estrazione del DNA da polvere d'osso.

PROBLEMATICHE DEL DNA DA POLVERE D'OSSO.

Problematiche:

- DNA in bassa quantità, frammentato e danneggiato
- Spesso contaminato da DNA di microorganismi
- Possibile presenza di composti organici ed inorganici del suolo (es. acidi umici, sali) inibitori nelle successive fasi di analisi

Esistono una serie di protocolli che vengono utilizzati per recuperare il DNA. Il principio fondamentale è l'utilizzo di **guanidina tiocianato o idrocloride** (leggi slide).

Comparison and optimization of ancient DNA extraction

Ancient DNA extraction from bones and teeth

Stefan Rohland & Michael Hofreiter

Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments

Guanidina tiocianato o idrocloride: sale caotropico che guida l'adsorbimento del solo DNA alle particelle di silice (fase solida). Agisce sequestrando gli inibitori presenti. Gli altri composti vengono lavati via, ed infine il DNA viene eluito.

LIBRERIA GENETICA: insieme di molecole di DNA in forma stabile che rappresenta il genoma (o porzioni di esso) di un organismo

Principio di preparazione dei campioni per l'NGS:

- Tutti i frammenti di DNA del campione vengono riparati alle estremità e legati ad adattatori universali. Gli adattatori permettono l'amplificazione e il sequenziamento di tutti i frammenti di DNA in parallelo usando dei primers universali.
- Esistono kit commerciali per la preparazione delle library, ma per il DNA antico l'efficienza con cui le molecole vengono riparate alle estremità e legate agli adattatori è spesso bassa.
- Sono stati quindi sviluppati dei protocolli per migliorare questa efficienza e monitorare la resa del campione al termine di ogni passaggio sperimentale.

Esistono due tipi di librerie:

- Double strand library (vista l'altra volta)
- Single strand library: costituiscono una libreria su entrambi i strand (forward e reverse) → vengono sequenziate entrambe → ho maggiore informazione perché le raddoppio.

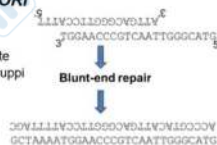
La libreria è costruita a partire da molecole di DNA a singolo filamento, in 3 step: 1) attacco del primo adattatore alle molecole a singolo filamento, mediante *T4 DNA ligasi*; 2) copia dei filamenti di DNA mediante *proofreading polymerase*; 3) attacco del secondo adattatore a doppio filamento mediante *T4 DNA ligasi*

- Molto più efficiente per DNA altamente degradato
- Molto più costosa della libreria double-stranded

Il genoma di Homo heidelbergensis è stato fatto con single strand library.

ATTACCO DEGLI ADATTATORI

Blunt-end repair: estremità sporgenti al 5' o 3' vengono pareggiate tramite ricostruzione della sequenza, mediante *T4 DNA polymerase*. Gruppi fosfati vengono attaccati al 5' mediante *T4 polynucleotide kinasi*



Adapter ligation: due adattatori (P5 e P7) vengono legati ad entrambe le estremità dei filamenti mediante *T4 DNA ligasi*



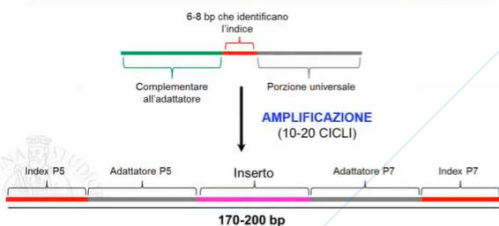
Adapter Fill-in: con *Bst polymerase* per rimuovere i nicks



qPCR (Real Time) con Solexa primer pair

ATTACCO DEGLI INDICI

Ogni campione viene marcato con una specifica coppia di indici → possibilità di sequenziare più campioni contemporaneamente nella stessa run → le sequenze prodotte vengono distribuite nei rispettivi campioni che le hanno generate grazie all'associazione INDICI-CAMPIONE



possiamo utilizzare UDG su una libreria:

PHILOSOPHICAL TRANSACTIONS B
rstb.royalsocietypublishing.org

Partial uracil-DNA-glycosylase treatment for screening of ancient DNA

Nadin Rohland^{1,2}, Eadaoin Harney^{1,2,3}, Swapnil Mallick^{1,2,3}, Susanne Nindorf^{1,2} and David Reich^{1,2,3}

URACIL-DNA GLICOSILASI (UDG) → enzima che riconosce l'Uracile e taglia il legame tra la base e lo zucchero creando un sito apiramidico

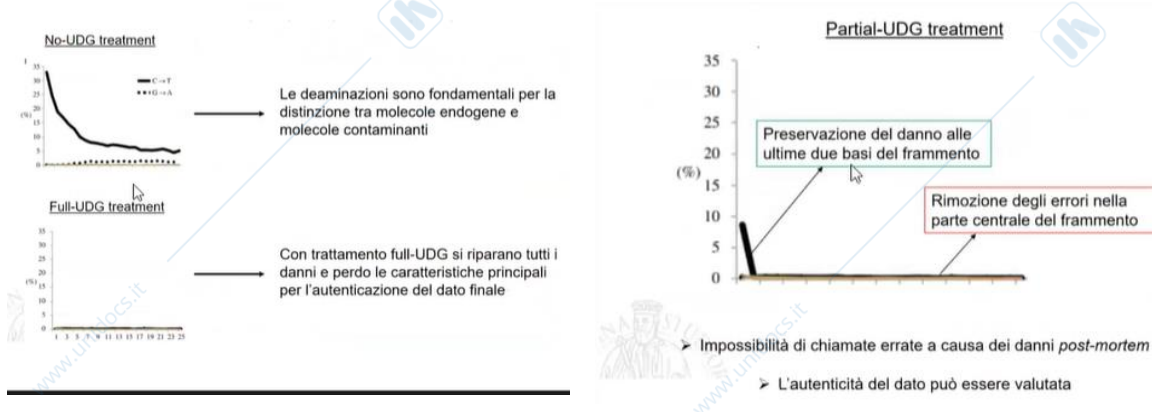
Nelle molecole degradate, l'UDG rimuove le citosine deaminate, mentre l'enzima ENDO-VIII taglia nei siti abasici risultanti riducendo i tassi di errore del DNA antico

→ elimino le C→T: è sbagliata utilizzarla perché non riconosco il DNA antico. Infatti non viene utilizzata completamente ma viene fatta facendo un'operazione "analisi con mezzo UDG" → ne utilizzo una quantità

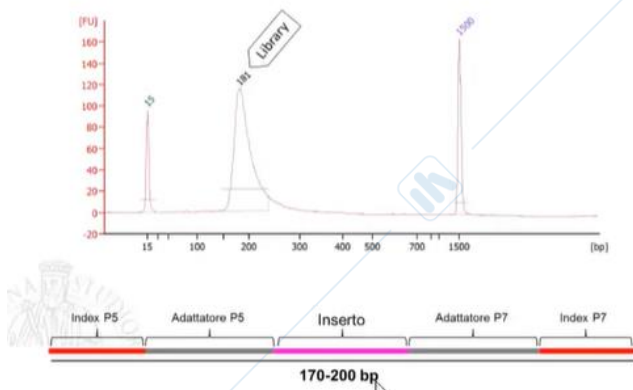
ridotta: quindi mi elimina le citosine all'interno del genoma, ma le lascia all'estremità 5' e 3' → per riconoscere il DNA antico perché ci permette di distinguere il DNA endogeno da esogeno.

UDG può essere utilizzato quando vado a fare una libreria per analizzare il genoma di un non umano → **Bos primigenius** perché non ho contaminazioni da parte di mucche moderne su bovini antichi.

Trattandosi di genomi umani: due possibilità: non lo utilizzo proprio o lo utilizzo a metà.



VISUALIZZAZIONE E INTERPRETAZIONE DEL RISULTATO



lunghezza frammento libreria: 170-200 bp in base alla lunghezza del frammento di DNA.

CATTURA DEL DNA MITOCONDRIALE CON TECNOLOGIA MARECHIC

OPEN ACCESS Freely available online
Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products
 Tomislav Marčić, Mark Whitten, Svante Pääbo

CATTURA: Metodologia che consente di selezionare, dal pool di molecole, solo frammenti di interesse mediante l'utilizzo di sonde specifiche.

La metodollogia marechic è simile alla PEC. Va a catturare il DNA mitocondriale, ma anche genomi nucleari. La differenza è la costruzione dei frammenti che vanno a catturare il DNA mit e più economico → molto meno costosa e meno laboriosa.

Il primer PEC costa molto e servono circa 500 primer PEC.

Preparazione di sonde andando a fare delle long range PCR su mitocondri attuali.

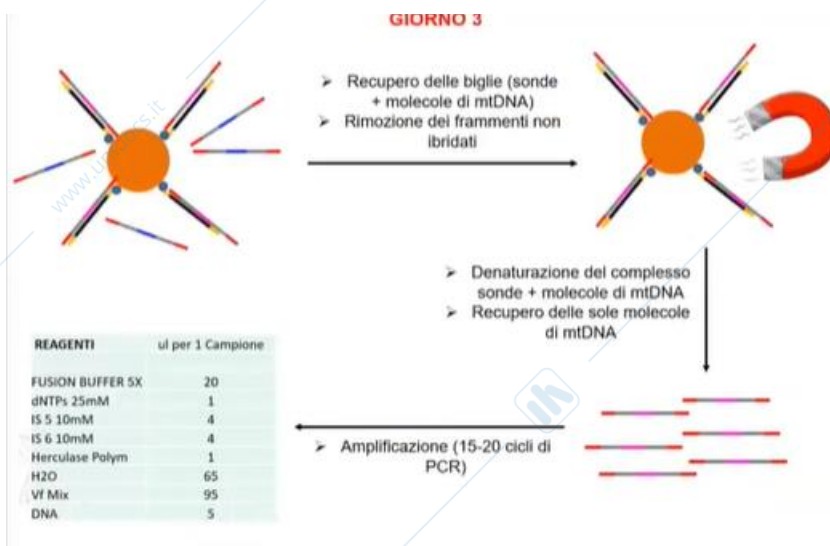
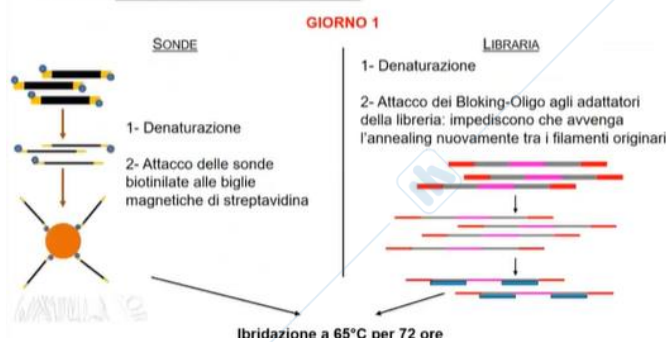
La metodollogia marechic si basa sulla costruzione di sonde mitocondriali andando ad amplificare con 2 long range PCR mitocondri umani. Fare due long range PCR → significa andare ad amplificare con due coppie di primer l'intero mitocondrio. → Dopo viene frammentato tramite sonificazione con lunghezza variabile → inserisco su questi frammenti di DNA mit delle molecole di biotina che a sua volta si legano a molecole di streptavina che stanno su biglie magnetiche. Queste biglie magnetiche vanno a reagire con una libreria che ho precedentemente preparato per la cattura dei frammenti di DNA che sono complementari → Dopo la cattura

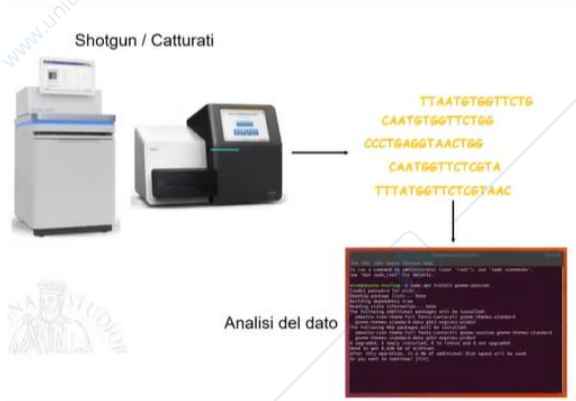
→ tramite un magnete rimuovo ciò che ho catturato dalla libreria → denaturo e recupero il DNA mitocondriale antico

È una metodica più semplice rispetto alla PEC, ma allo stesso tempo mi può portare a grossi risultati perché ho la capacità di catturare molti frammenti grazie alla ridondanza di sonde che ho ottenuto dalla frammentazione del DNA mitocondriale che ho precedentemente amplificato e poi biotinizzato ec..



-> 500 ng totali di sonde (250 ng Mt1 + 250 ng Mt2)
 -> 2000 ng totali di Libreria (si catturano 2-3 campioni in una stessa miscela di reazione combinati in maniera equimolare)





LEZIONE 16 NOVEMBRE

Con la tecnologia marelchic possono essere catturati anche frammenti di DNA nucleare.

Ci sono diversi modi per procedere a questo tipo di cattura.

In un articolo mostra alcuni esempi di come è possibile utilizzare delle catture sia in liquido sia in solido con delle sonde complementari al DNA da catturare.

CATTURA FRAMMENTI/GENI DI DNA NUCLEARI

ARTICOLO: ANCIENT DNA STUDIES: NEW PERSPECTIVES ON OLD SAMPLES.
(vedi)

- **Cattura in solido:**

Cattura su un SeqCap di microarray: Se vogliamo catturare tutti quei geni che presentano delle mutazioni puntiformi o SNPs che ci permettono di determinare caratteristiche fenotipiche di quell'individuo

Funzionamento di flusso di lavoro su genoma nucleare

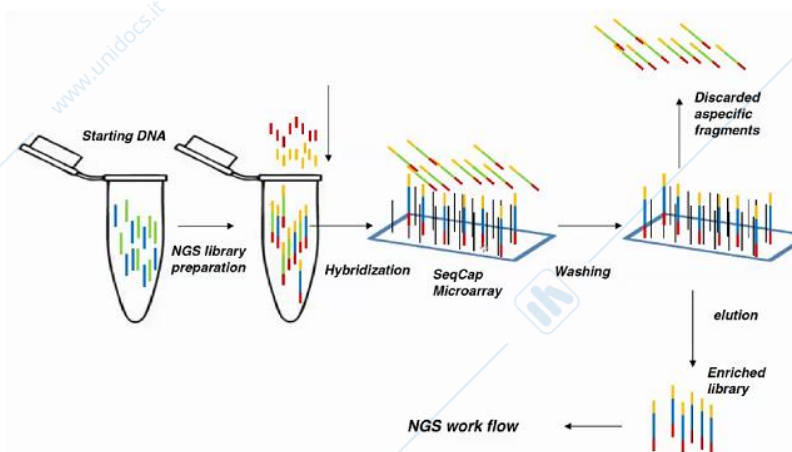


Figure 4 Nimblegen sequence capture workflow. In solid-sequence, capture is performed by hybridization between NGS library sequences and the capture probes immobilized onto an array surface.

DNA di partenza nucleare → preparazione libreria → ibridare la libreria con un microarray: superficie in silice dove ho predisposto delle sequenze complementari alle regioni che voglio andare a catturare → così vado a catturare ciò che successivamente voglio andare a leggere. → eluzione per eliminare i frammenti specifici → arricchire la libreria (amplifico) → NGS work flow.

- **Cattura in liquido**

DNA di partenza → preparazione libreria → sonde biotate a RNA complementari a regioni di DNA che voglio catturare con la libreria che ho preparato → avrò complessi sonda + libreria → eliminazione di ciò che non ho catturato → digestione RNA → mi rimane DNA che ho catturato → amplificazione → NGS

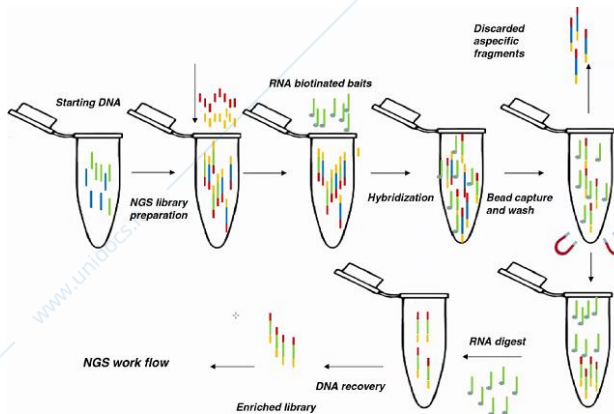


Figure 5 Agilent SureSelect sequence capture workflow. In solid-sequence, capture is performed by hybridization between NGS library

Associato a ciò ci sono analisi di carattere bioinformatico che ci permettono di ottenere dei risultati attendibili attraverso il processamento che tutte questi frammenti che ho catturato → ottengo delle informazioni che mi permettono di capire se il campione è autentico o no.

In base a cosa si sceglie la cattura in solido o in liquido? In base alla resa che voglio ottenere.

Nel caso di una cattura in solido: utilizzo dei chip che vengono vendute con specifiche caratteristiche, cioè con numero di sonde specifico.

In una cattura in liquido possono costruire io le sonde e quindi catturare ciò che voglio.

APPLICAZIONE ANTROPOLOGICA su campioni di POMPEI

è un sito famosissimo, oltre ad avere una serie suggestioni di carattere archeologico, è una città ben preservata → Si possono ricostruire un tratto genetico di città romana del 79 DC → dovuto all'eruzione del vesuvio che ha cristallizzato la struttura di questa popolazione: le abitudini (ciò che mangiavano, come si vestivano) ec.. → si possono estrarre molto informazioni sia di vita quotidiana, ma anche un tratto genetico della città. → E' un progetto nel lab del prof: the genetic portrait of Pompei.

Ricostruzione genetica su alcuni individui rinvenuti a pompeii in contesti particolari.

Giuseppe Fiorelli, archeologo di Pompei decise, laddove vedeva delle possibilità di trovare un corpo sotto la cenere, decise di attuare un processo che permetteva di ricostruire con un calco di gesso l'individuo che si trovava sotto lo strato di cenere. Sotto il calco in gesso sono conservate le ossa → Recupero di informazioni genetiche dalle ossa che erano al di sotto di questi calchi

Cosa è successo nel 79 d.C

*hanno datato i campioni con carbonio → conferma l'eruzione nel 79 d.C anche si è confuso con il periodo perché è stato tra fine settembre inizio ottobre.

Testimonianza del plinio il giovane.

SCAVI ARCHEOLOGICI: il contributo di Giuseppe Fiorelli

"A Pompei ci si tuffa nella vertigine dei millenni dove l'illusione del tempo è caduta e l'umanità si disvela."
Margherita Sarfatti, 1924



Tecnica del calco in gesso



TECNICA DEL CALCO IN GESSO:

intende quella **tecnica** utilizzata negli scavi archeologici vesuviani mediante la quale, tramite **gesso** o cemento e acqua, è stato possibile recuperare la forma di esseri umani e animali e oggetti vegetali vittime dell'eruzione **del Vesuvio del 79** → Consiste nel buttare il gesso sopra il corpo ottenendo il calco in gesso con sotto le ossa.

I primi individui sono stati analizzati nella casa del bracciale d'oro (la casa si chiama così perché uno degli individui aveva un anello d'oro) → 4 individui: 2 grandi e 2 bambini → è stata ribattezzata la famiglia della casa del bracciale d'oro in cui questi individui avrebbero raccontato un dramma di quello che poteva essere successo a queste persone durante l'eruzione.

I campioni sono stati nominati.



gli archeologi hanno chiesto se si trattava di una famiglia: hanno analizzato il DNA mitocondriale → ottengo un'uguaglianza di genomi tra figli e madre.

Altri calchi: nominate le fanciulle che si abbracciano → hanno chiesto di capire se erano imparentate: alla fine è stato visto che erano due uomini e non erano imparentati

NGS WORKFLOW

ESTRAZIONE DNA → COSTRUZIONE LIBRERIA → CATTURA DNA MIT CON TECNOLOGIA MARCHIC → SEQUENZIAMENTO CON ILLUMINA

Il DNA si degrada quando ci sono alte temperature → com'è possibile avere DNA in buono stato di conservazione? Le ossa sono rimaste incombuste

Ci sono delle metodiche: Quantifier Trio DNA quantification kit che ci permette di capire in base al rapporto tra la quantità piccola di DNA autosomico e quantità grande di DNA autosomico quanto è conservato il nostro DNA. Si può calcolare un indice di degradazione che ci dice la possibilità di ottenere la possibilità di poter studiare il DNA e non era male.

In tutti i campioni c'era Y target → in tutti i campioni c'erano molecole prodotte dal nostro quantifier.. che matchavano su delle molecole di cromosoma Y su questi campioni, tranne SKC-32.

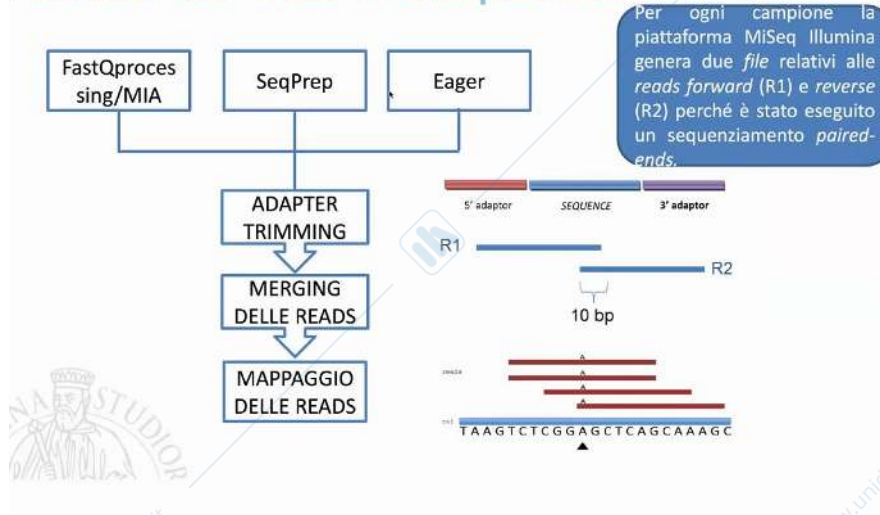
INDICE DI DEGRADAZIONE

Small autosomal

Large autosomal

Sample	Concentrazione DNA (pg/μl)		
	Large autosomal target	Small autosomal target	Y target
PC-33	0	2,9	2,8
PC-34	6,1	117	107
SKC-21	0	28,1	22,9
SKC-23	4,3	32,7	25,4
SKC-32	NON DETERMINATO		
SKC-35	1,2	17,8	9,3

Tabella 1: Risultati della quantificazione degli estratti con il Quantifier Trio DNA Quantification Kit. I valori di concentrazione per ciascun campione, espressi in picogrammi (pg/μl), sono relativi alle regioni target selezionate: Large Autosomal, Small Autosomal e Y-Chromosome.

**Analisi dei dati di sequenza**

RISULTATI erano buoni → hanno consentito di andare avanti

Parametri

Average coverage: media di copertura per ciascun campione → quante volte è stata ciascuna base di DNA mit.

Average length: Lunghezza media dei frammenti che sono riuscito a recuperare

Covarege fold: quante volte leggo il Dna mit. Covarege 1 fold 99,87 (coperto tutto il mitocondrio una volta) significa che ho bene il campione una volta → cioè ho letto tutte le basi del mio DNA mit.

AUTENTICITA' DELLE SEQUENZE:

tramite analisi dei pattern di misincorporazione con **MAP damage** in 5'→3'. → risultati ottimali → ci dicono che sono autentiche

Analisi del *pattern* di contaminazioni

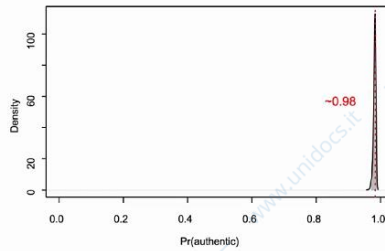
CONTAMIX: confronta la sequenza consenso dei campioni con le sequenze di 311 profili mitocondriali che rappresentano potenziali contaminanti.

Sample	Map authentic
PC-33	0,858146
PC-34	0,903813
SKC-21	0,954469
SKC-23	0,982695
SKC-32	0,998837
SKC-35	0,876246



Tabella 7: valore di map authentic ottenuto dalla stima del pattern di misincorporazione con il software ContamMix.

SCHMUTZI: il livello di contaminazione del campione viene stimato sulla base dell'analisi del *pattern* di deaminazione (*ContDeam*) e dell'allineamento delle sequenze del campione con le sequenze di 197 individui europei rappresentati da potenziali contaminanti (*mtCont*).



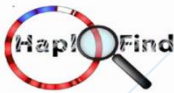
altre analisi per verificare autenticità:

Contamix confronta la sequenza consenso dei campioni che abbiamo ottenuto con sequenze di 311 profili mitocondriali che rappresentano potenziali contaminanti → ci da informazioni della probabilità che un mitocondrio appartenga ad una fonte biologica, quindi non è muscuglio di piu DNA mit. Se è superiore al 90% → è una

sequenza che non è nata da fenomeni di contaminazioni, ma è tutto DNA endogeno.

Potrebbe essere uno strumento da utilizzare in forense → Es.in situazioni di omicidi, dove si ricerca DNA mit e se non si capisce molto dal DNA mit si utilizza questo si capisce che appartiene ad un individuo o piu individui diff.

Determinazione dell'aplogruppo mitocondriale



HAPLOFIND: software online che confrontando la sequenza consenso del campione con la sequenza di riferimento rCRS ne individua i polimorfismi assegnando l'aplogruppo di appartenenza.

Sample	Haplogroup
PC-33	N1b1a1
PC-34	U1a1
SKC-21	H1h1
SKC-23	H
SKC-35	T2n

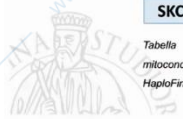
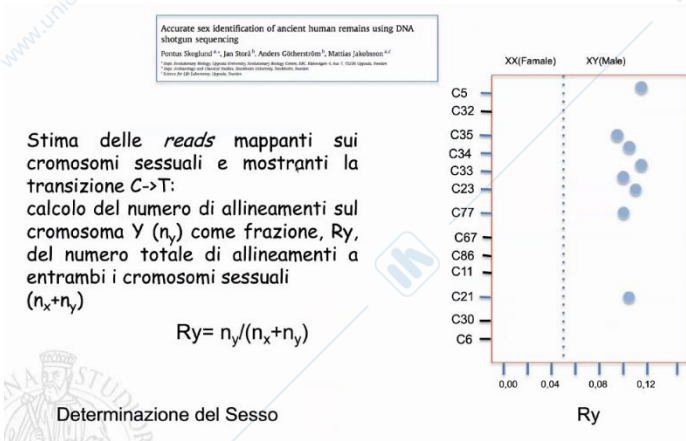


Tabella 8: identificazione degli aplogruppi mitocondriali ottenuta con il software on line HaploFind.

Software: **HAPLOFIND** per la determinazione dell'aplogruppo mitocondriale → sono stati ottenuti mitocondri differenti → E' stato scoperto che la famiglia della casa del bracciale d'oro non erano imparentati, ma erano tutti maschi.

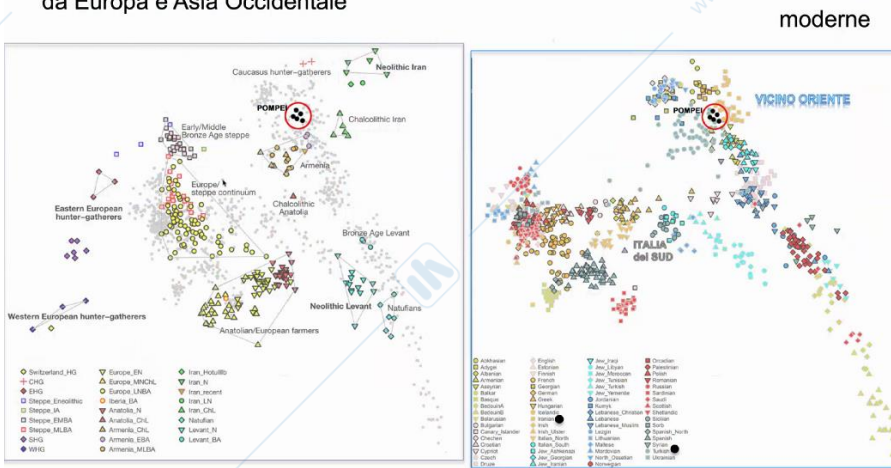
Lo stesso per le fanciulle, sono due individui maschi e non sono imparentati tra di loro.



Dall'analisi del genoma nucleare si è ottenuto lo stesso risultato per la determinazione del sesso → si trattava di tutti maschi.

ANALISI DELLE COMPONENTI PRINCIPALI (PCA)

- Rappresentazione grafica delle frequenze alleliche negli individui di Pompei in relazione ad un *database* di circa 900 individui moderni e oltre 200 individui antichi da Europa e Asia Occidentale



è stata fatta anche un'analisi sul genoma nucleare con la metodica di cattura di frammenti DNA nucleare con sonde biotinate che indicavano caratteristiche nucleari di questi individui → è stato visto che degli individui di pompei erano più vicini alle popolazioni del mediooriente → probabilmente erano dei servitori.

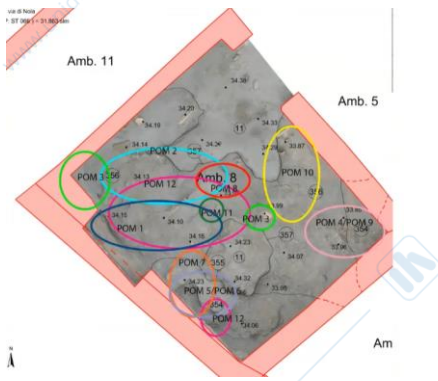
E' stata fatta anche analisi di carattere isotopico su smalto dentale: ci dice che questi individui hanno trascorso una parte di vita nel medioriente e una parte a pompei. Ipotesi è che fossero dei servitori

LEZIONE 23 NOVEMBRE

“LA STANZA DEGLI SCHELETRI”

Racconta come gli scavi dello scorso anno sono state ritrovate in una stanza diversi reperti scheletrici: volevano vedere se c'erano relazioni di carattere parentale.

CERCHI evidenziati



Indicano i riferimenti che gli archeologi hanno dato ai vari ritrovamenti

- Il colore indica l'idea del ritrovamento
- la grandezza del cerchio ci dice le possibili connessioni sia dal punto di vista anatomico e parentale degli individui presenti all'interno della stanza.

Es: l'archeologo chiede: POM2 e POM12 sono due frammenti ossei diversi che appartengono allo stesso individuo?

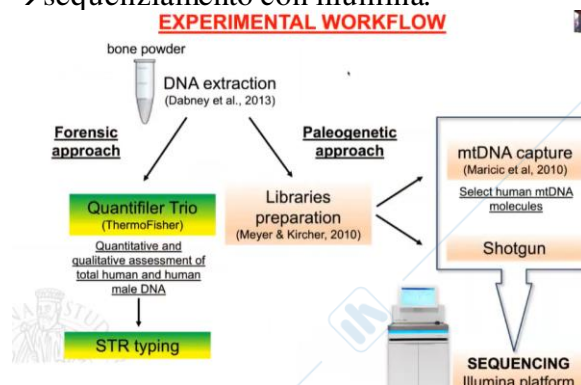
Su alcuni frammenti scheletrici c'è scritto che sono adulti, in altri non sono sicuri

1) **Il campionamento:** Viene fatto il campionamento di rocche petrose e denti → recupero polvere d'osso.

2) estrazione DNA

Poi si può eseguire

- un approccio forense: quantifier trio che permette di verificare la quantità di DNA in buono stato di conservazione e di verificare il sesso e STR typing per stabilire le relazioni parentali.
- un approccio paleogenetico (preparazione libreria, cattura DNA mit → shotgun → sequenziamento con illumina.



FLUSSO DI LAVORO PER IL SHOTGUN LIGHT

Ci permette di visualizzare la possibilità di recupero di DNA in buon stato di conservazione → significa di sequenziare tutto il genoma, ma non di leggerlo a fondo (cioè di leggerlo molte volte) ma di avere un'idea della conservazione del DNA all'interno del reperto.

Attraverso lo shotgun light → si può determinare anche il sesso via molecolare, che abbiamo visto che si determina anche con la metodologia classica

1) **Preparazione librerie:** double-stranded library → Ci si approccia quando siamo sicuri di ottenere una buona resa di DNA su un determinato reperto. Questa valutazione viene fatta in base all'antichità del reperto. I reperti di Pompei non sono molto antichi, e anche se recuperati in contesti molto particolari (forme di calore molto elevata).

ATTACCO DEGLI ADATTATORI

Blunt-end repair: estremità sporgenti al 5' o 3' vengono pareggiate tramite ricostruzione della sequenza, mediante T4 DNA polymerase. Gruppi fosfati vengono attaccati al 5' mediante T4 polynucleotide kinasi

Blunt-end repair

Adapter ligation: due adattatori (P5 e P7) vengono legati ad entrambe le estremità dei filamenti mediante T4 DNA ligasi

Adapter Fill-in: con Bst polymerase per rimuovere i nicks

qPCR (Real Time) con Solexa primer pair

Blunt end repair è importante perché ci permette di andare a vedere le misincorporazioni in 5' - 3' → per distinguere DNA endogeno da esogeno.

Dopo l'attacco degli adattatori si passa all'indexing PCR → inserimento indici per fare un pool di lettura.

Agilent Bioanalyzer 2100

Reagenti:

- Gel-dye mix -> matrice all'interno del quale avviene la separazione dei frammenti di DNA, contenente l'intercalante fluorescente
- DNA ladder-> standard esterno; miscela di frammenti di peso molecolare noto, usato come riferimento per la valutazione dei campioni
- DNA Marker -> standard interno; allineamento dei dati del Ladder con quelli del campione

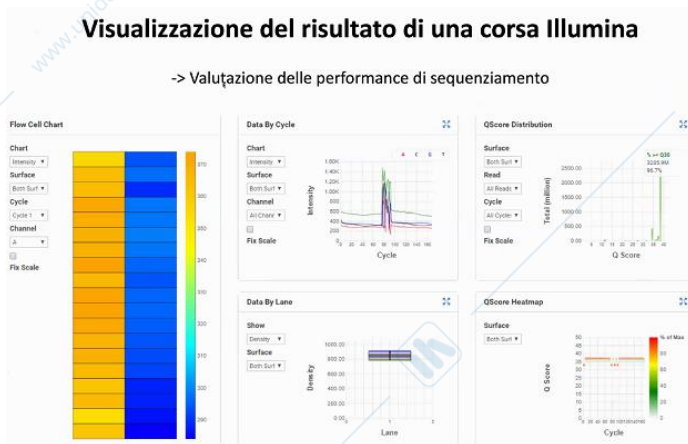
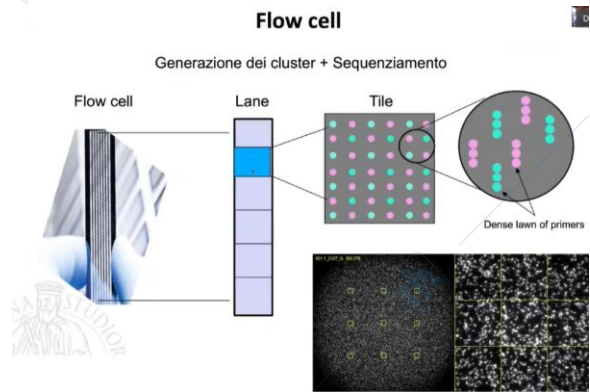
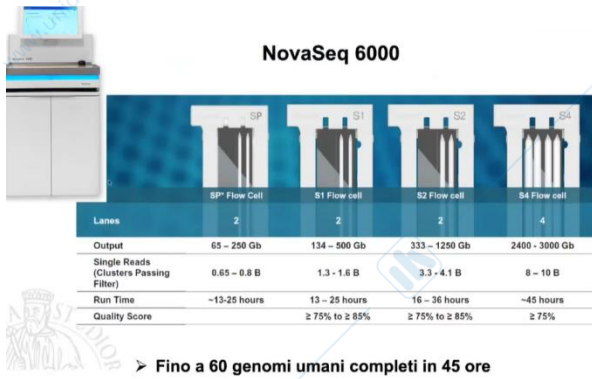
marker + Campioni (5 uL) (1 uL)

Gel-dye (9 uL)

Marker + Ladder (5 uL) (1 uL)

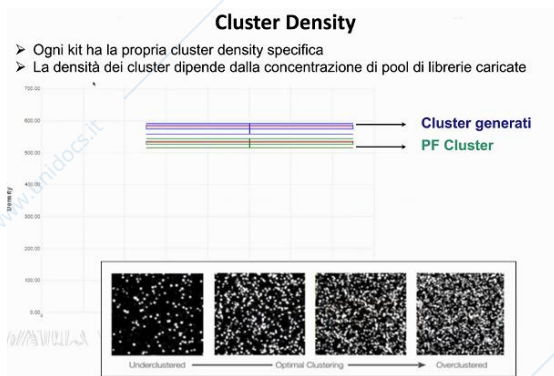
Attraverso Agilent Bioanalyzer abbiamo la possibilità di andare a misurare se la libreria è stata fatta bene o no. È come se fosse un elettroforesi → ci dice la lunghezza dei frammenti della libreria

SEQUENZIAMENTO: nel caso di Shotgun light non si va a catturare il mitocondrio, ma si passa direttamente al sequenziamento della libreria con NovaSeq che sequenza i campioni su cui è stata fatta la libreria (fino a 3 tera di sequenziamento) (fino a 60 genomi umani in 45 ore) 9P0se



Ogni kit (S1,S2,S3,S4) ha la propria cluster density specifica.

La densità dei cluster dipende dalla concentrazione di pool di librerie caricate.



Per shotgun light si intende un sequenziamento leggero (light sequencing) → 5 milioni di reads per ciascun campione

Perché si fa il sequenziamento light?

• **Per uno screening iniziale del campione** → si misura

➤ %DNA endogeno: consente di stimare quanto è il DNA umano sul tot delle reads ottenute → quindi di stabilire la preservazione del materiale genetico del campione. *Numero tot delle read ottenute/ numero di read che mappano sul genoma umano*

Attraverso BLAST che permette di vedere le sequenze di DNA che abbiamo ottenuto che combaciano con DNA umano.

- Tasso di misincorporazione
- Pattern di frammentazione: la lunghezza media dei frammenti di DNA antico varia tra 40 e 150 bp, essendo un DNA altamente degradato.

→ indicano che se il DNA mappante è effettivamente antico o deriva da contaminazioni moderne

Le molecole antiche = alta frequenza di CoT

Le molecole moderne → contaminanti = bassa frequenza di CoT

- Nt/mt ratio: molecole mitocondriali/ molecole nucleari → mi indica se possono estendere la stima delle contaminazioni ottenuta sul mtDNA al nuDNA. (copertura media sul mitocondrio/copertura media sui cromosomi) → abbiamo la possibilità di capire la qualità del DNA endogeno.

MT/NT RATIO (rapporto molecole mitocondriali/molecole nucleari) → mi indica se posso estendere la stima delle contaminazioni ottenuta sul mtDNA al nuDNA

$$\frac{\text{Copertura media sul mitocondrio}}{\text{Copertura media sui cromosomi}} = \text{Mt/nt ratio}$$

infatti all'interno della rocca si trova DNA in buono stato di conservazione

❖ Il rapporto tra DNA mitocondriale e nucleare (mt/nc) è correlato negativamente con un aumento del contenuto di DNA endogeno



- ❖ Il valore varia in base al distretto osseo analizzato:
 Rocca → Mt/nt ~ 100
 Denti → Mt/nt ~ 80
 Post-craniale → Mt/nt ≤ 200

Quindi attraverso lo screening del campione iniziale si può passare al sequenziamento del genoma completo o a catturare SNP nucleari di nostro interesse.

• **Determinare il sesso** in maniera accurata → applicazione importante dal punto di vista evolutivo ma soprattutto archeologico.

➤ DETERMINAZIONE BIOLOGICA DEL SESSO

Accurate sex identification of ancient human remains using DNA shotgun sequencing
 Pernicelli Soglund¹, A. B. Jen Stork², Anders Götherström³, Martin Jakobsson^{4*}

❖ Identificazione del sesso considerando il numero reads che si mappano sui cromosomi X e Y

$$R_Y = n_Y / (n_X + n_Y)$$

(n_Y) = number of alignments to the Y chromosome
 (n_X) = number of alignments to the X chromosome

R_Y ~ 0.0022



R_Y ~ 0.09



Questo metodo richiede almeno 100.000 sequenze mappanti sul genoma umano → non sempre si hanno questi numeri, in particolare per resti antichi mal conservati.

considerando il numero di reads che mappano sul cromosoma X e sul cromosoma Y.

Per avere un valore del rapporto accettabile dal punto di vista statistico dobbiamo avere almeno 100.000 sequenze mappanti sul genoma umano.

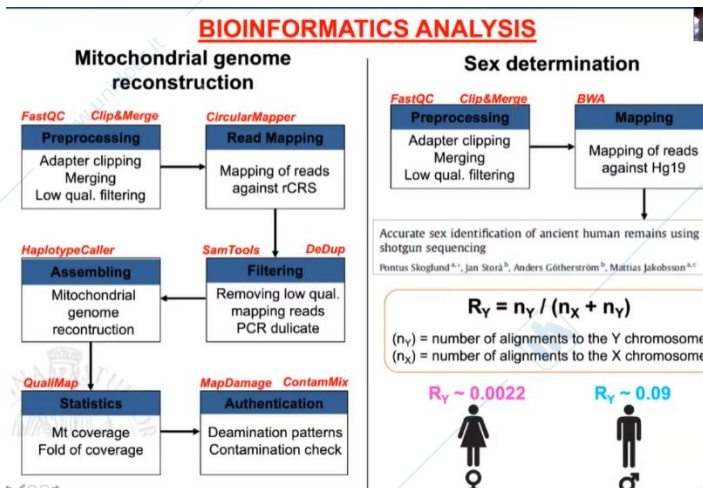
Qualora si abbia difficoltà ad ottenere un numero di reads mappanti elevato

Per la determinazione del sesso c'è anche un approccio ulteriore, anche se meno preciso
numero di reads mappanti sul cromosoma X / copertura autosomica
 quando si utilizza?? Quando si ha quantitativi di DNA minori.

Senza la conoscenza di queste metodiche non si possono ricavare informazioni, come nel caso specifico di Pompei associano questi studi a studi di carattere antropologico, di carattere genetico di popolazioni. Anche per ricostruire le dinamiche evolutive della nostra specie, queste metodiche ci permettono di fare ciò.

*perché proprio Homo? Queste metodiche possono essere applicati ad individui che sono vissuti 400-500 mila anni fa. Il più antico è Homo del Belgesis.

Ritornando su Pompei



Step bionformatici per arrivare alla validazione del dato.

- 1) Ricostruzione genoma mitocondriali (legge l'immagine)
- 2) determinazione del sesso

Ciò è stato fatto sugli individui di pompei: verificando i DNA mitocondriali per verificare se ci sono relazioni parentali.

è stato fatto anche un approccio di carattere forense andando a vedere i profili nucleari STR.

È stato visto anche %DNA endogeno → valori molto buoni.

Conclusion

- 1) Sex Determinations on 10 samples
- 2) According to mtDNA analysis two group maternal related
- 3) STR data confirms the mt DNA relationship
- 4) Good amount of endogenous DNA in several samples

In the next future

Deep shotgun sequencing in order to confirm the kinship and other information form the genome

LEZIONE 25.11

STR: COSA SONO E COME SI ANALIZZANO

Su alcuni campioni di pompeii sono stati anche fatti gli **STR**

Gli **STR** o short tandem repeat → Sono regioni non codificanti del DNA extragenico.

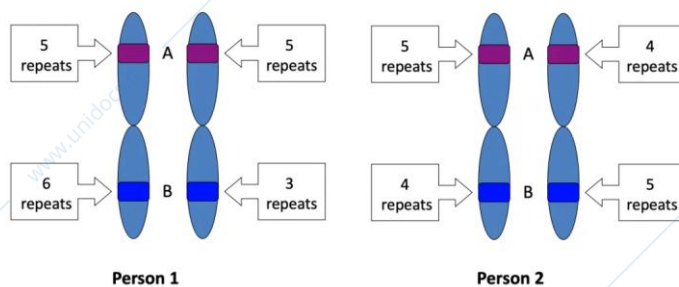
Sono marcatori fondamentali per gli studi di antropologia forense e l'identificazione personale, anche in studi di genetica di popolazione, per studiare la variabilità genetica di un popolazione ecc..

L'utilizzo che tratteremo sarà quello in antropologia forense.

GLI STR appartengono ai polimorfismi di lunghezza: Può essere in eterozigosi e monozigosi:

Viene analizzato la lunghezza di questi STR.

Polimorfismi di lunghezza



Gli STRs teranucleotidici

In genetica forense la principale tipologia di STRs impiegate sono le ripetizioni tetranucleotidiche e, talora, le ripetizioni pentanucleotidiche in tandem

discriminante è basso

In genetica forense le principali STR impiegate sono le **ripetizioni tetranucleotidiche** e, talora le ripetizioni **pentanucleotidiche** in tandem → Perché essendo **rare e più variabili** ci permettono → *un'identificazione più certa e più semplice*, quindi il potere discriminante è elevato. Rispetto a quelle più frequenti, dove il potere

Introduzione alla biologia forense: basi molecolari dell'identificazione genetica

Tipologie e caratteristiche degli STRs tetranucleotidici

Nella seguente tabella sono indicati: il **tipo di STR**, la **localizzazione cromosomica**, il **core**, gli **intervalli (range) possibili delle dimensioni in bp dei singoli STR** ed il **tipo di fluorocromo coniugato ad ognuno di essi**

STR system	Chromosomes	Common Sequence Motif	Range (bp)	Fluorophores
D3S1358	3p	TCTA(TCTG) ₁₋₃ (TCTA) _n	114-142	5-FAM
vWA	12p12-pter	TCTA(TCTG) ₃₋₄ (TCTA) _n	157-197	5-FAM
FGA	4q28	(TTTC) ₃ TTTTTCT(CTTT) _n C TCC(TTCC) ₂	219-267	5-FAM
Amelogenina	X: p22.1-22.3 Y: p11.2	/	107-113	JOE
TH01	11p15.5	(AATG) _n	169-189	JOE
TPOX	2p23-2per	(AATG) _n	218-242	JOE
CSF1PO	5q33.3-34	(AGAT) _n	281-317	JOE
D5S818	5q21-31	(AGAT) _n	135-171	NED
D13S317	13q22-31	(GATA) _n	206-234	NED
D7S820	7q	(GATA) _n	258-294	NED

Considero STR system D3S1358 e il locus amilogenina.

Il D3S1358 si trova sul braccio p del cromosoma 3 → vediamo quanto è lungo il frammento → Range bp da 114 a 142 → significa che ci saranno alleli che avranno una lunghezza in questo range. → in base alla ripetizioni n che sono presenti su quel tipo di motivo.

Ora consideriamo il gene omologo dell'amilogenina

→ 106-112: 106 è sul cromosoma X, 112 cromosoma y

Quando utilizziamo i pannelli STR (significa amplificazione contemporanea di tutti i loci nel STR system) → si ha una PCR multiplex → amplifico un numero superiore a 1 di loci (in questo caso 10) → su un individuo avrò dai 10 ai 20 prodotti di amplificazione: 10 prodotti di amplificazione → perché tutti i prodotti di ampl su ciascun locus avranno la stessa lunghezza → quindi individuo è omozigote su ciascun locus.

È probabile che l'individuo possa essere eterozigote per tutti i loci → 20 prodotti di amplificazione (Un individuo ad un locus può essere omozigote o eterozigote)

Se è omozigote → se è amplifico su entrambi i cromosomi → prodotto avrà la stessa lunghezza.

Se avrà la stessa lunghezza se lo faccio correre in un campo elettroforetico → avrò una banda.

Se il mio individuo è omozigote a D3S1358, vWA FGA, amilogenina (quindi è femmina) ecc.. a tutti i loci → avrò 10 prodotti di amplificazione)

Esempio: se ho al D3S1358 un individuo omozigote con un range allelico di 114 e un individuo maschio con range allelico 107-113 significa che il frammento 114 arriverà in contemporanea con il frammento di 113 bp → Domanda: se questi due frammenti 114 e 113 arrivano in contemporanea vedo o il frammento di 114 o 113.

Mi faccio aiutare dalla colonna accanto dove c'è scritto fluorofori → perché questo sistema il D3S1358 → ha come fluoroforo 5-FAM → è legato alla coppia di primer che amplifica D3S1358

Guardando il gene omologo dell'amilogenina che viene amplificato con una coppia differente di primer che ha attaccato un altro fluoroforo JOE che emette una luce differente → quindi quando arrivano contemporaneamente i prodotti di amplificazione di D3S1358 e del gene dell'amilogenina e mettendo una luce laser che manda un segnale laser che mi riflette il colore del fluoroforo → distinguerò qual è l'allele che contiene il fluoroforo 5-FAM e l'allele che contiene il fluoroforo JOE?? SI! → **Le PCR multiplex fatte per identificare i profili STR avranno primer marcati con fluorofori differenti tutte le volte in cui i profili allelici di un determinato sistema STR si**

sorrappongono con i profili allelici di un altro sistema STR → in questo modo posso riconoscere gli STR di un locus rispetto ad un altro.

Introduzione alla biologia forense: basi molecolari dell'identificazione genetica

Tipologie e caratteristiche degli STRs tetranucleotidici

Tecnicamente gli STR, a confronto con altri marcatori, presentano il vantaggio di essere:

1. Numerosi
2. Equamente distribuiti in tutta la parte eucromatica del genoma
3. Altamente polimorfici e quindi informativi
4. Analizzabili facilmente tramite PCR
5. Basso tasso di mutazione → bassa probabilità di cambiamenti da una generazione all'altra

Ultimo punto è importante perchè se gli STR sono utilizzati nell'identificazione personale e nelle relazioni parentali (padre-figlio), se ci fossero grossi cambiamenti tra una generazione e l'altra non potrebbero essere utilizzati per le relazioni parentali.

Tassi medi di mutazione per vari polimorfismi (per locus per generazione)

- VNTR $10^{-1} - 10^{-2}$
- STR $10^{-2} - 10^{-4}$
- SNPs $10^{-6} - 10^{-8}$
- Indel (retrovirus) $10^{-10} - 10^{-11}$
- Nella regione ipervariabile del DNA mitocondriale, valori fino a 5×10^{-5} per sito per generazione

Introduzione alla biologia forense: basi molecolari dell'identificazione genetica

Tipologie e caratteristiche degli STRs tetranucleotidici

- **Small product sizes** are generally compatible with degraded DNA and PCR enables recovery of information from small amounts of material → $\leq 1\text{ng}$ of DNA to type 13-15 STR loci
- **Multiplex amplification** with fluorescence detection enables high power of discrimination in a single test
- Power of discrimination ranges from 10^{14} - 10^{23} . World population is 10^9 so bring on the database!
- Commercially available in an easy to use kit format
- Uniform set of core STR loci provide capability for national and international sharing of criminal DNA profiles

Le multiplex PCR di loci STRs

- La tecnica elettiva di analisi dei polimorfismi STRs del DNA è l'amplificazione tramite PCR di più regioni contemporaneamente (multiplex PCR) e la successiva separazione e rilevazione degli ampliconi tramite elettroforesi capillare (EC)
- La coamplificazione e la rilevazione di più loci STRs è consentita dalla possibilità di distinguere i diversi ampliconi generati in una reazione di multiplex:
 - ✓ Sulla base della sola M, - Ampliconi di M, differente, e quindi distinguibili tramite la sola determinazione della M, vengono marcati con gli stessi fluorofori
 - ✓ Sulla base della M, e della differente marcatura con fluorofori - ampliconi di M, analoga, e quindi non distinguibili tramite la sola determinazione della M,, vengono marcati con fluorofori differenti

sono commercialmente utilizzati in kit.

Sunto analisi STR.

*Mr: lunghezza.

Le strategie analitiche: l'amplificazione

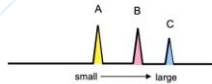
Le multiplex PCR di loci STRs

(A) Simultanea amplificazione di tre regioni sul DNA stampo



(B) Separazione e rilevazione dei prodotti di PCR tramite:

- M, dell'amplicone
- il tipo di marcatura fluorescente dell'amplicone



Kit utilizzati per i loci STR: (da ricordare questo): gli STR vengono venduti in KIT commerciali che vanno da un numero minimo di 9 a 17 loci in co-amplificazione.


Le strategie analitiche: l'amplificazione

Le multiplex PCR di loci STRs

 **U.S. Core Loci** : CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Amelogenin

 **UK/European Core Loci (ESS)**: FGA, TH01, VWA, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, Amelogenin
Recommended Loci: D1S1656, D2S441, D10S1248, D12S391, D22S1045, TPOX

 **German Core Loci**: FGA, TH01, SE33, VWA, D3S1358, D8S1179, D18S51, D21S11, Amelogenin

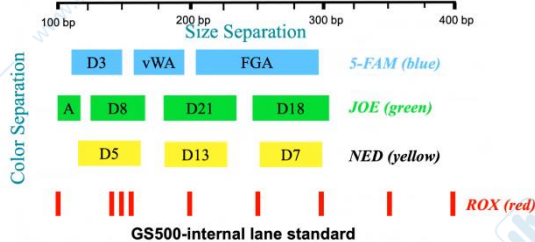
 **Interpol Standard Set of Loci (ISSOL)**: FGA, TH01, VWA, D3S1358, D8S1179, D18S51, D21S11; Optional: Amelogenin

Uno dei KIT più utilizzati: AmpFISTR

Le strategie analitiche: l'amplificazione

Le multiplex PCR di loci STRs AmpFISTR® Profiler Plus™

Kit available from PE Biosystems (Foster City, CA)



utilizzo la stessa colorazione quando la loro lunghezza in termini di polimorfismo per quel determinato STR è differente quindi non si sovrappongono mai.

Su DNA altamente degradato → utilizzo STR è complicato → quindi occorrono frammenti ossei (rocche petrose) che restituiscono DNA in buono stato di conservazione.

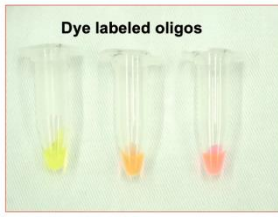
Le strategie analitiche: l'amplificazione tramite PCR

Le multiplex PCR di loci STRs

- La marcatura degli ampliconi è consentita dal legame covalente di uno dei due primers, progettati esterni alla sequenza ripetuta ed utilizzati per amplificare le sequenze dei microsatelliti, con differenti fluorofori
- Tali molecole assorbono l'energia fornita da un laser per poi riemetterla a lunghezze d'onda differenti specifiche per ciascun fluoroforo

Le strategie analitiche: l'amplificazione

Il sistema di marcatura fluorescente

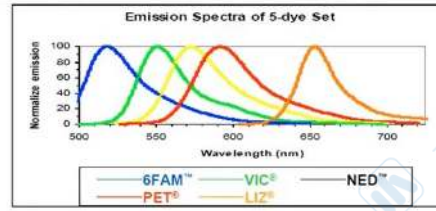


6FAM (yellow), VIC (orange), NED (red)

- UV Spec to determine concentration
- HPLC to evaluate purity
- TOF-MS to confirm correct sequence
- CE (ABI 3130xl) to determine presence of residual dye molecules ("dye blobs")

Le strategie analitiche: l'amplificazione

Il sistema di marcatura fluorescente

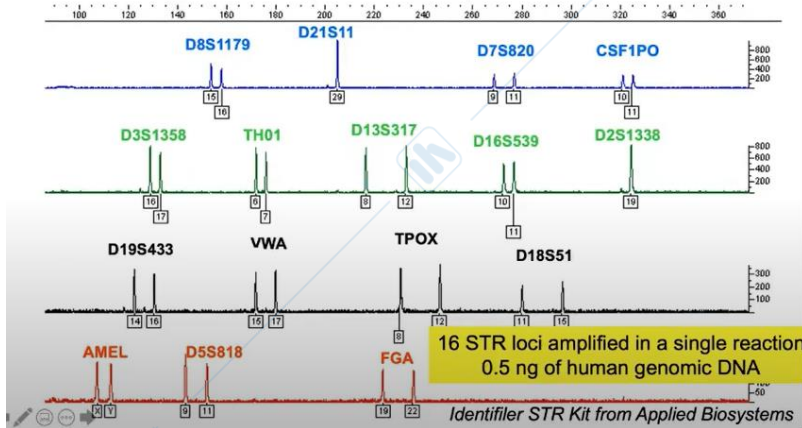


- AmpFtSTR® Identifier® PCR Amplification Kit
- AmpFtSTR® Yfiler® PCR Amplification Kit
- AmpFtSTR® SEfiler™ PCR Amplification Kit
- AmpFtSTR® MiniFiler™ PCR Amplification Kit

Per visualizzare i fluorofori si fa un elettroforesi capillare e ottengo un cromatogramma

Le multiplex PCR di loci STRs

Information is tied together with multiplex PCR and data analysis



Profilo STR per 16 loci.

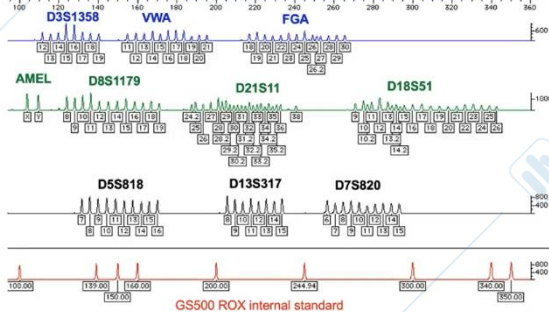
Es. D8S1179: due picchi (15 e 16) → è eterozigote per quel locus.

Il D21S11 è omozigote.

Questo individuo è maschio perché presenta due picchi nel gene dell'amlogenina.

La multiplex PCR di loci STRs

Profiler Plus kit allelic ladders (Applied Biosystems)



varianti alleliche per ogni determinato locus.

LEZIONE 30 NOVEMBRE

IMPIEGO DEI POLIMORFISMI DEL CROMOSOMA Y IN ANTROPOLOGIA FORENSE

Il cromosoma Y è un marcatore che permette di fare analisi di carattere filogenetico, popolazionistico e antropologico-forense (per riconoscimento individuale, per riconoscimento tra padre-figlio, tra fratelli) → ereditarietà di carattere unilineare da parte paterna.

È uno strumento relativamente recente per quanto concerne il suo studio. Studi che hanno avuto inizio nel 2003.

La storia recente del cromosoma Y

- "Full" Y-chromosome sequence became available in June 2003; over 200 Y-STR loci identified (only ~20 in 2000)
- Selection of core Y-STR loci (SWGDAM Jan 2003)
- Multiple commercial Y-STR kits released
 - Y-PLEX 6,5,12 (2001-03), PowerPlex Y (9/03), Yfiler (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

È uno strumento utilizzato anche in studi di antropologia forense: Uno dei più frequenti è l'analisi di paternità e in casi di stupro.

Caratteristiche strutturali e funzionali

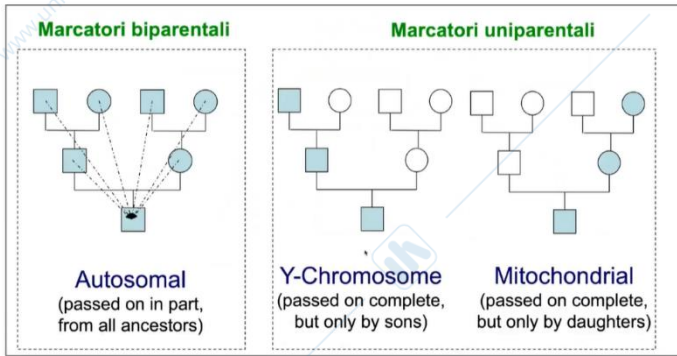
- Il cromosoma Y umano è un piccolo cromosoma acrocentrico costituito da circa 60 Mbp
- possiede pochissimi geni funzionali (78 contro i ~2000 del cromosoma X), alcuni presenti anche sul cromosoma X, altri Y specifici
- per lo più geneticamente inerte, costituito da eterocromatina costitutiva composta da differenti tipi di DNA non codificante
- scambi di materiale genetico per crossing-over avvengono all'interno di piccole regioni di omologia note come regioni pseudoautosomiche, poiché non mostrano una tipica eredità legata al sesso; esse sono denominate PAR1 (~2.6 Mbp) e PAR2 (<1 Mbp)

PAR 1 e PAR 2 sono agli estremi del cromosoma.

number	PAR	number
Transcription factor - sex determination 1 <i>SRXY</i>	1	<i>RPS4Y</i> 1 Protein of small ribosomal subunit
Testis transcript 1 m <i>TTY1</i>	2	<i>ZFY</i> 1 Zinc finger transcription factor
Cyclin B binding protein m <i>TSPY</i>	3	<i>PCDH11</i> 1 Protocadherin - cell adhesion
Protein tyrosine phosphatase m <i>PRY</i>	4A	<i>PRKY</i> 1 Ser/Thr protein kinase
Testis transcript 1 m <i>TTY1</i>	4B	<i>AMELY</i> 1 Tooth enamel formation
Testis transcript 2 m <i>TTY2</i>		Centromere
Cyclin B binding protein m <i>TSPY</i>		
	5	<i>USP9Y</i> 1 Deubiquitinating enzyme
		<i>DBY</i> 1 DEAD-box - RNA helicase
		<i>UTY</i> 1 TPR-motif
		<i>TB4Y</i> 1 Actin sequestration
		<i>VCY</i> 2 Variable charged protein
Chromodomain protein m <i>CDY</i>		<i>SMCY</i> 1 Transcription factor
Membrane transport protein m <i>XKRY</i>		<i>EIF1AY</i> 1 Translation initiation factor
		<i>RBMY</i> 30 RNA-binding protein
Protein tyrosine phosphatase m <i>PRY</i>	6	<i>RBMY</i> 30 RNA-binding protein
Testis transcript 2 m <i>TTY2</i>		
RNA-binding protein 4 <i>DAZ</i>		
Basic protein m <i>BPY2</i>		
Protein tyrosine phosphatase m <i>PRY</i>		
Chromodomain protein m <i>CDY</i>		
Y-chromosome genes not found on the X	7	Heterochromatin
	PAR	Y-chromosome genes with homologs on the X

Il cromosoma Y, essendo un frammento di DNA trasmesso in maniera unilineare → può essere ascritto a quella categoria di marcatori → **APLOTIPI DEL CROMOSOMA Y**, come il DNA mitocondriale.

Viene utilizzato come **MARCATORI UNIPARENTALI**, grazie ad una grossa parte non ricombinante, come il DNA mit. → è possibile andare a ritroso nel tempo per ricostruire l'ereditarietà del cromosoma Y → permette di ricostruire una genalogia.



nei marcatori di carattere autosomico, attraverso i fenomeni di ricombinazione impediscono di andare a ritroso nel tempo

mtDNA
 SNPs (Single Nucleotide Polymorphism)
 • sequenza HVR
 • RFLPs regione codificante

cromosoma Y
 SNPs (Single Nucleotide Polymorphism)
 STRs (Short Tandem Repeats)

Il DNA mit presenta solo polimorfismi di sequenza.

Il cromosoma Y presenta polimorfismi di sequenza e lunghezza.

L'insieme dei polimorfismi che sono associati fisicamente tra di loro su una molecola di mtDNA e sul cromosoma Y definisce un **APLOTIPO**

I marcatori uniparentali: mtDNA e cromosoma Y

Aplotipo mtDNA
 la somma dei siti delle variazioni della sequenza nucleotidica

Aplotipo	16038	16126	16187	16189	16223	16264	16270	16278	16293	16300	16311
rif.	A	T	C	T	C	C	C	C	A	A	T
1	G	C	T	C	T	T	T	T	G	T	C
2		C	T	C	T	T	T	T	G	C	
3		C	T	C	T	T	T	T		C	
4		C	T	C	T	T	T	T		C	

Aplotipo cromosoma Y
 la somma della variabilità di polimorfismi STR

Aplotipo	DYS19	DYS389-1	DYS389-2	DYS390	DYS391
1	15	8	25	21	10
2	15	10	28	21	10
3	17	10	27	21	10
4	16	10	27	21	10

Allo stesso modo, possono essere tracciati alberi filogenetici che definiscono il DNA mit, ma anche il cromosoma Y con i suoi aplogruppi.

I marcatori uniparentali: mtDNA e cromosoma Y

Aplogruppo mtDNA
 • Si definisce sulla base della condivisione di mutazioni specifiche in posizioni con un basso tasso di mutazione (stabili)
 • Aplotipi non identici possono appartenere allo stesso aplogruppo

Aplogruppo Y
 Si definisce sulla base della condivisione di mutazioni specifiche per marcatori biallelici (SNPs), e non per i microsatelliti (tasso di mutazione troppo alto)



APPLICAZIONI CROMOSOMA Y:

- ANALISI FORENSE
- ANALISI GENEALOGICO
- STUDIO EVUZIONISTICO

Impiego in genetica forense: caratteristiche

In genetica forense lo studio dei polimorfismi del cromosoma Y presenta alcune peculiari caratteristiche dovute a:

1. natura uniparentale olandrica di tali marcatori
→ trasmessi inalterati (a meno di mutazioni) lungo la discendenza maschile
2. assenza di ricombinazione intercromosomica
→ no riassortimento dei marcatori da una generazione alla successiva

genetica forense

Impiego in antropologia forense: vantaggi

- ✓ I polimorfismi del cromosoma Y sono presenti solo su tale cromosoma, mentre polimorfismi omologhi non sono presenti sul cromosoma X → possono essere analizzati solo nei maschi (che, comunque, commettono circa il 98% dei crimini più violenti), escludendo il DNA femminile (amplificazione per PCR maschio-specifica)
- ✓ Tutti i membri maschi di una medesima discendenza presentano il medesimo aplotipo Y a meno di mutazioni → nella ricerca di persone scomparse, vittime da disastri di massa (DVI), in paternità controverse e in studi genealogici è possibile utilizzare come campione di confronto qualunque soggetto maschile della medesima discendenza
- ✓ Attualmente analizzabili, con elevata attendibilità, fino a 17 regioni STRs del cromosoma Y (aplotipo di 17 STRs)
- ✓ Linee guida internazionali per l'analisi e l'interpretazione del dato analitico definite

Impiego in antropologia forense: vantaggi

- ✓ Metodi di analisi analoghi a quelli impiegati nell'analisi dei polimorfismi STRs autosomici con un limite di rilevazione maggiore e un'interpretazione del dato (profilo ad allele singolo) più semplice
- ✓ Impiego prioritario nel contesto dei reati di violenza sessuale perché solo la componente maschile viene rilevata (senza il bisogno di effettuare una estrazione differenziale del DNA) → nei casi di miscele genetiche tra due o più soggetti la risoluzione di tali miscele è semplificata (soprattutto determinazione n° contributori)
 - Violenze sessuali da parte di maschi azospermici o vasectomizzati (non è presente sperma per estrazione differenziale del DNA)
 - Violenze sessuali di gruppo (almeno per esclusione)
 - Recupero del profilo del DNA dell'aggressore esteso fino ad oltre le 48 h
 - Miscele genetiche di vari fluidi biologici maschio-femmina e maschio-maschio

Con l'utilizzo dell'STR maschile è possibile andare a recuperare il DNA maschile e differenziarlo dal DNA femminile a livello degli STR in quanto sono presenti solo sul maschio.

Essendo un marcatore unilineare → si osserva un picco → perché è in fase aploide

Impiego in ANTROPOLOGIA forense: limitazioni

- ✓ Eredità patrilineare → tutti i membri maschi di una medesima discendenza presentano il medesimo aplotipo Y a meno di eventi mutazionali → in casi di violenza sessuale e, comunque, in tutti i contesti in cui è stato commesso un crimine, sulla base dell'aplotipo Y-STRs non è possibile escludere a priori un'intera discendenza maschile
- ✓ I loci Y-STRs sono concatenati l'uno all'altro (per assenza di ricombinazione) a costituire un aplotipo che può essere considerato come un singolo locus multiallelico (altamente polimorfico), quindi non sono trasmessi in maniera indipendente l'uno dall'altro → il potere di discriminazione è estremamente inferiore a quello di un profilo genetico autosomico e la RMP non può essere calcolata

se ci sono delle piccole mutazioni non è possibile accusare una persona se il suo cromosoma Y presenta delle differenze rispetto a quelle trovate sulla scena del crimine.

Impiego in ANTROPOLOGIA forense: limitazioni

- ✓ Sono necessari studi popolazionistici per valutare la diversità degli aplotipi e le loro frequenze nella popolazione
- ✓ Possibilità di considerare un aplotipo Y indipendente rispetto ai loci STRs autosomici (?)

per capire se questo STR è discriminante devo vedere la frequenza nella popolazione → se tutti gli individui hanno la stessa variabilità per quel determinato STR non è discriminante.

L'impiego del cromosoma Y in ANTROPOLOGIA forense

Tipologie di marcatori polimorfici Y

- **STRs (microsatellites)**
 - DYS19, DYS385, etc.
 - mostly tetranucleotide repeats
- **Bi-allelic markers (unique event polymorphisms--UEP)**
 - SNPs (single nucleotide polymorphisms)
 - Y *A_u* polymorphism (YAP) or other insertions/deletions ("indels")
- **Minisatellite**
 - MSY1 (DYF155S1) composed of 48-114 copies of a 25 bp repeat unit with 5 sequence variant repeat types
 - typed by MVR-PCR (minisatellite variant repeat)

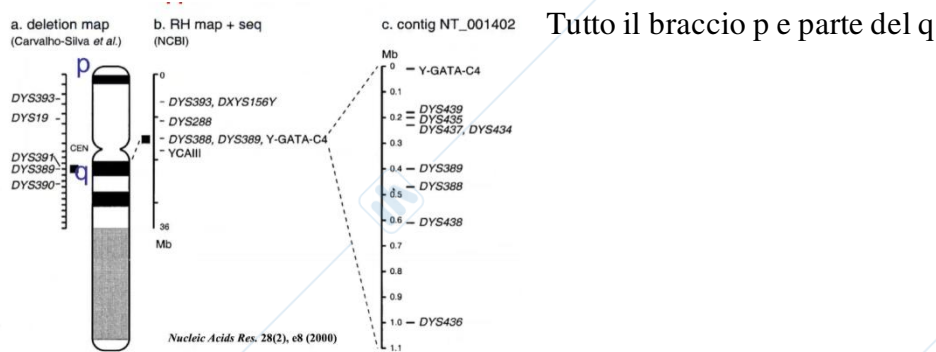
L'impiego del cromosoma Y in ANTROPOLOGIA forense

I marcatori STRs del cromosoma Y

Marker Name	GenBank Accession	Repeat Motif	Allele Range	PCR Product Sizes	Reference
DYS19	X77751	TAGA	8-16	178-210 bp	Roewer 1992
DYS385	Z93950	GAAA	10-22	252-300 bp	
DYS388	G09695	ATT	12-17	128-143 bp	
DYS389 I	G09600	(TCTG) (TCTA)	I: 7-13	239-263 bp	
DYS389 II	G09600	(TCTG) (TCTA)	II: 23-31	353-385 bp	
DYS390	G09611	(TCTA) (TCTG)	18-27	191-227 bp	
DYS391	G09613	TCTA	8-13	275-295 bp	
DYS392	G09867	TAT	7-16	236-263 bp	
DYS393	G09601	AGAT	9-15	108-132 bp	
YCAIII	AC008370	CA	19-25	192-204 bp	Kayser 1997
DYS434	AC002992	ATCT	8-11	110-122 bp	Ayub 2000
DYS435	AC002992	TGGA	9-13	210-228 bp	Ayub 2000
DYS436	AC005820	GTT	10-15	128-143 bp	Ayub 2000
DYS437	AC002992	TCTA	8-11	186-202 bp	Ayub 2000
DYS438	AC002531	TTTTTC	6-12	203-233 bp	Ayub 2000
DYS439	AC002992	AGAT	9-14	238-258 bp	Ayub 2000
Y-GATA-A4	G42670	AGAT	11-14	242-254 bp	White 1999
Y-GATA-A7.1	G42675	ATAG	7-12	161-181 bp	White 1999
Y-GATA-A7.2	G42671	TAGA	8-12	174-190 bp	White 1999
Y-GATA-A8	G42672	TCTA	8-14	219-244 bp	White 1999
Y-GATA-A10	G42674	TATC	11-14	160-172 bp	White 1999
Y-GATA-C4	G42673	TATC	11-16	251-271 bp	White 1999
Y-GATA-H4	G42676	TAGA	10-13	362-370 bp	White 1999

Most Commonly Used Markers

LOCALIZZAZIONE STR SUL CROMOSOMA Y



L'interpretazione del dato analitico

Le valutazioni

Ottenuto un profilo (aplotipo) Y-STRs si deve:

1. Verificarne la qualità in termini analitici, in accordo con linee guida internazionali;
2. Calcolarne la frequenza di comparsa in database popolazionistici di riferimento;
3. Effettuare comparazioni e riferirne gli esiti

Le valutazioni biostatistiche: la stima della frequenza di un aplotipo

Si deve calcolare la frequenza con cui l'aplotipo Y-STRs estrapolato si trova in quella particolare popolazione di riferimento (X/N) mediante il metodo della conta (analogo a quello che si fa per il mtDNA) → numero di volte che un aplotipo è osservato in un database popolazionistico.

$$\text{Frequenza aplotipo} = \frac{X}{N}$$

X=numero di volte in cui l'aplotipo compare
N=numero di campionaria del database)

Per capire che probabilità c'è, dato un individuo con quel determinato STR, che quell'STR sia = ad un altro per effetto del caso oppure è proprio il ricercato e lo si dice in base a quanto è frequente.

Se la frequenza è alta → il potere discriminatorio è basso

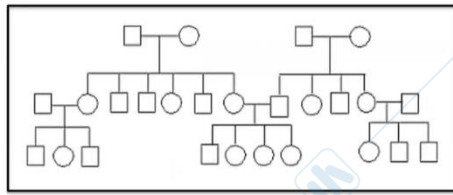
Se la frequenza è bassissima → potere discriminatorio è alto.

Alcuni STR amplificano frammenti molto grandi, ma siccome non sempre si lavora con DNA ben conservato → sono stati anche realizzati i **mini-STR** → che amplificano piccoli frammenti STR, in modo da essere utilizzati in contesti di DNA altamente degradato.

PROGETTO LEONARDO

È uno studio di carattere antropologico che riguarda la genealogia di Leonardo da Vinci: voleva verificare se le ossa conservate nella cripta in Francia appartenevano a Leonardo

Albero genealogico



□ maschi
○ femmine

ALBERO GENEALOGICO DEL CROMOSOMA Y:
possiamo fare la genealogia sugli SNP o sugli STR dell'Y.

Avere gli Y dei maschi dell'ultima generazione, dei discendenti di Leonardo da Vinci, se sono tutti uguali tra di loro → significa avere l'Y di Leonardo da Vinci

Per realizzare questo progetto occorre andare a ricercare i discendenti attuali e occorre andare a trovare DNA antico da tombe in cui sappiamo essere sepolti discendenti ai tempi di Leonardo → campione sia gli uni e sia gli altri.

Fare analisi sul cromosoma Y di un individuo attuale (fare la PCR), è molto semplice, mentre su quelli vissuti nel passato occorre un approccio NGS + target enrichment capture → sono state sviluppate sonde per catturare tutto il cromosoma Y per catturare ca 79 mila SNP con 41,145 sonde (simile alla tecnica Marecic) → Sequenziamento attraverso il MAPPING della sequenza del cromosoma Y come riferimento : per fare analisi su persone che sono oggi in vita, e sulle persone che non sono più in vita per confrontare le sequenze del cromosoma Y → In questo modo se si osserva la stessa sequenza su genealogia differenti attraverso la conoscenza del tipo di ereditarietà di questo cromosoma → saremo in grado di ricostruire il cromosoma Y di Leonardo.

A questo punto si può verificare se le reliquie hanno lo stesso cromosoma Y di Leonardo → se è la stessa sequenza → è stato dimostrato e si può analizzare il genoma completo

STR su campioni antichi: dove e come possiamo fare gli STR su campioni degradati.

Il DNA viene estratto da rocche petrose: in questo modo restituisce ben conservato: studio condotto su 13 individui longobardi di cui si sapevano le relazioni parentali, datati dal 6 al 7 secolo C.E (dopo Cristo) e sono stati testati attraverso gli STR.

Su un campione antico, andando ad estrarre DNA dal dente o dal femore il numero di loci che si riesce ad utilizzare è nullo o inferiore rispetto alla rocca petrosa.

(risenti)

*possiamo fare relazioni parentali s campioni antichi non utilizzando STR quando non abbiamo rocche petrose ma utilizzando SNPs

LEZIONE 2 DICEMBRE

ANALISI BIOINFORMATICA SU DNA ANTICO

Ci permettono di poter discriminare tra un DNA endogeno ed esogeno

Ci sono vari tipi di macchine che sequenziano il DNA: miniSEQ e Iseq (NovaSEQ versione aggiornata di hiSEQ → che permette di leggere fino a 60 genomi umani)

	MiniSeq System	MISeq Series	NextSeq Series	HiSeq Series	HiSeq X Series*
Key Methods	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing	Small genome, amplicon, and targeted gene panel sequencing	Everyday exome, transcriptome, and targeted resequencing	Production-scale genome, exome, transcriptome sequencing, and more	Population- and production-scale whole-genome sequencing
Maximum Output	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb
Maximum Reads per Run	25 million	25 million†	400 million	5 billion	6 billion
Maximum Read Length	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4-24 hours	4-55 hours	12-30 hours	<1-3.5 days (HiSeq 3000/HiSeq 4000) 7 hours-6 days (HiSeq 2500)	<3 days

NovaSeq 6000				
	S1 Flow Cell	S1 Flow Cell	S2 Flow Cell	S4 Flow Cell
Lanes	2	2	2	4
Output	65 - 250 Gb	134 - 500 Gb	333 - 1250 Gb	2400 - 3000 Gb
Single Reads (Clusters Passing Filter)	0.65 - 0.8 B	1.3 - 1.6 B	3.3 - 4.1 B	8 - 10 B
Run Time	~13-25 hours	13 - 25 hours	16 - 36 hours	~45 hours
Quality Score	≥ 75% to ≥ 85%	≥ 75% to ≥ 85%	≥ 75% to ≥ 85%	≥ 75%
➤ Fino a 60 genomi umani completi in 45 ore				

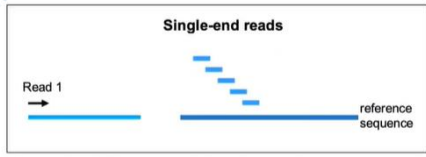
1) preparazione libreria 2) Bridge PCR: perché le molecole si posizionano a ponte, si forma il cluster, cioè un insieme di molecole clonali tutte identiche l'una con l'altra

3) Sequenziamento: sequenziamento si svolge a partire da un vetrino: noi utilizzeremo il sequenziamento su piattaforma ILLUMINA -> per questo tipo di sequenziamento si utilizza un vetrino chiamato **FLOW-CELL**: è proprio come un vetrino da microscopio ed è la piattaforma su cui avvengono tutte le reazioni necessarie allo svolgimento del sequenziamento. La FLOW-CELL, a seconda del modello di sequenziatore che si usa, può essere: divisa in tanti sotto-canalini chiamati **LANE**. Quindi l'alternativa non è un vetrino suddiviso, ma è un vetrino liscio e unico. Queste sotto divisioni si chiamano **LANE** e una lane a sua volta è suddivisa in altre sotto regioni che si chiamano **TILE**: il tile altro non è che una regione di spazio dove sono ancorati dei primer di sequenziamento complementari agli adattatori che abbiamo attaccato alle sequenze e sono quelli che danno l'avvio sia alla bridge PCR, sia al sequenziamento vero e proprio. Il **TILE** inoltre è anche quella piccolissima regione di spazio in cui viene rilevata l'immagine dalla **CCD CAMERA**: perché durante l'avvio del sequenziamento si ha il rilascio di picchi luminosi in base al tipo di fluoroforo e quindi in base al nucleotide che viene incorporato e questo picco luminoso viene rilevato da una foto. Il tile è il pezzettino che di volta in volta viene fotografato e scansionato. Ovviamente, come per le foto, maggiore è il numero di **TILE** e migliore sarà la risoluzione dell'immagine, per cui più accurato sarà il sequenziamento.

Le lane inoltre possono essere costituite o da 13 o da 19 **TILE**: di solito quello che utilizziamo è quella da 19 tile in modo tale da avere un'altissima risoluzione dell'immagine e quindi un'ottima lettura.

COME AVVIENE IL SEQUENZIAMENTO?

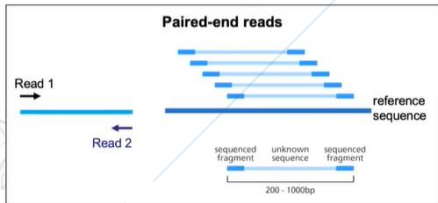
È un sequenziamento che avviene per sintesi utilizzando 4 nucleotidi, ciascuno dei quali è marcato con un fluoroforo di colore diverso; questi nucleotidi, avendo il fluoroforo, una volta incorporati bloccano la sintesi, per cui quando mando il primo ciclo si attaccano i fluorofori alla regione corrispondente, ma poi la sintesi non può procedere perché il fluoroforo legato impedisce l'attacco del nucleotide successivo. Per cui faccio partire il primo ciclo, si attaccano i nucleotidi alla base corrispondente e a questo punto viene identificata dalla **CCD camera** l'immagine, viene cioè fatta una foto dell'immagine e successivamente viene associato a ciascun colore, la base corrispondente. Il sequenziamento inoltre può essere di 2 tipi diversi



SINGLED-END SEQUENCING: sequenziamento solo in forward: 1 read per frammento di DNA

E' il metodo più semplice ed economico per utilizzare il sequenziamento Illumina

SINGLE-END SEQUENCING viene usato per sequenziare genomi attuali.



PAIRED-END SEQUENCING: il frammento di DNA viene sequenziato sia in forward che in reverse: 2 reads per frammento di DNA

Il sequenziamento in paired-end consente un miglior allineamento delle reads e quindi un'identificazione più accurata dei riarrangiamenti genetici come inserzioni e delezioni

PAIRED-END SEQUENCING viene fatto quando abbiamo a che fare su campioni di DNA degradato

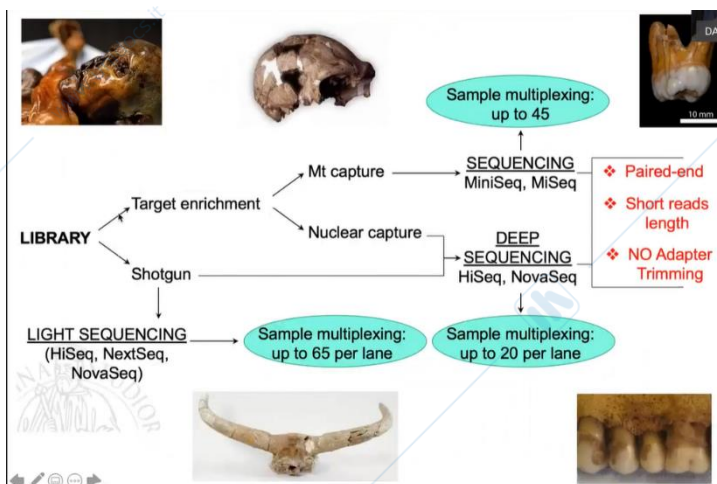
Vengono descritti nel sequenziatore anche i dati per ciclo, i dati per LANE, altri risultati come Qscore distribution (numero di reads che vengono lette) e la superficie dove questo avviene.

È importante anche la **densità di cluster**

- underclustered: poche generazioni di sequenziamento su un cluster → da poche informazioni
- Optimal clustering: in cui il cluster in cui avviene il sequenziamento non è molto affollato da read → lettura ottimale
- Overclustered: ci sono molte read → non riusciamo a distinguere una read dall'altra.

La densità di cluster dipende dal pool di libreria caricata.

Flusso di lavoro su DNA degradato da campioni antichi



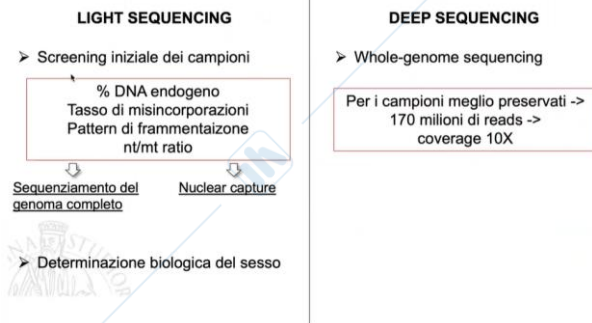
Quindi in base alla tipologia di studi che vogliamo fare, si utilizza un sequenziamento piuttosto che un altro.

Per l'analisi sul genoma mitocondriale, visto che è piccolo, si utilizza il MY SEC perché grazie alle sue 25 milioni di reads generate, è più che sufficiente per ottenere le letture relative ai mitocondri. Se invece si deve mappare il nucleare si utilizza o il MY SEC o l'HIGH SEC perché sul nucleare abbiamo bisogno di una

maggiore potenza di sequenziamento *Novaseq e Hiseq hanno un output più elevato →

perciò lo si usa per capture nu

SHOTGUN SEQUENCING



Determinazione biologica del sesso lo si fa quando:

- campioni infanti o subalduti, anche se c'è il bacino.
- manca parti anatomiche come il bacino negli adulti.

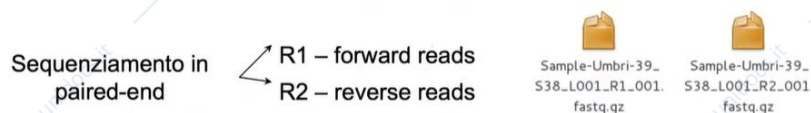
Con l'avvento delle NGS la situazione si è complicata: prima di tutto perché devo gestire una gigantesca mole di dati che non è più una sola lettura per ciascun campione, ma sono migliaia di letture per ciascun campione e ci sono numerosi campioni per LANE (grazie agli indici che ci permette di distinguere una libreria da un'altra). Quindi la primissima fase, indispensabile prima di iniziare la ricostruzione del frammento è:

EFFETTUARE IL DE-MULTIPLEXING è un attività di carattere bioinformatico che separa i vari campioni attraverso la lettura degli indici --> viene letto il nome del campione e gli indici associati a questo campione, per cui, facendo questa associazione, **NOME DEL CAMPIONE-INDICI**, lui riesce ad associarmi a ciascun campione le reads che portano quegli indici → viene fatta da due software che si trovano all'interno dello strumento, quindi questa parte qui la fa ancora il sequenziatore. I software che il sequenziatore utilizza sono **CASABA** e, quello che si chiama il **REPORTER SOFTWARE** (nel caso del My Seq) e **blc2fastq.pl**

Lo strumento fa le varie associazioni e crea dei file: sono file OUTPUT in formato **FAST-Q: file di testo** dove vengono separate le reads che hanno tra loro indice uguale.

Ho anche bisogno del **SIMPLE SHEET**: ha un formato EXCEL, è una tabella in cui nella prima parte ci sono le informazioni legate agli esperimenti che stiamo facendo, per cui c'è il nome dell'operatore, il progetto, la data di sequenziamento ecc. poi c'è una sezione in cui è riportata l'informazione legata alle reads, cioè alle letture che voglio ottenere; per cui ci sono scritti il numero dei cicli utilizzati e il numero degli indici, cioè 2, e gli indici per ogni campione.

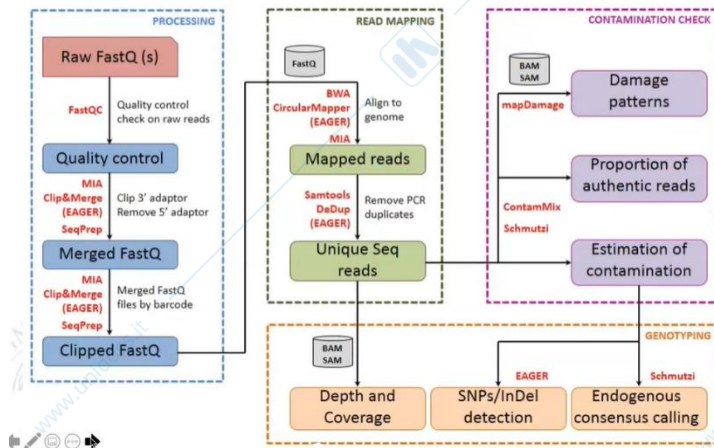
Per i campioni antichi si fa il sequenziamento **PAIRED-END**: Per ciascun campione ottengo due file: uno è quello dove sono elencate tutte le sequenze in forward e l'altro è quello dove ci sono tutte le sequenze in reverse e sono quelle che si chiamano R1 e R2,



Una volta ottenuti questi file li devo trattare in modo tale che alla fine possa ricostruire solo ed esclusivamente la sequenza consenso del campione

WORK FLOW GENERALE BIOINFORMATICO SU CAMPIONI ANTICHI

WORKFLOW GENERALE



Si può dividere in 4 parti:

- 1) Processing
- 2) Read mapping
- 3) Contamination check
- 4) Genotyping

Ci sono quindi una serie di processi per ottenere un genoma ricostruito.

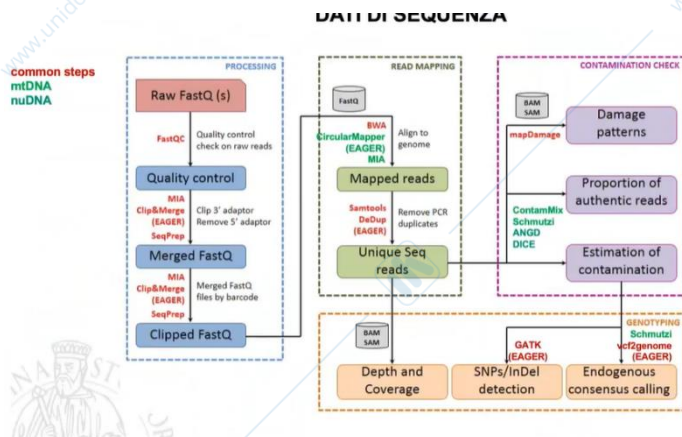
Entrando nel dettaglio del PROCESSING

Lettura FastQ che derivano dal de-multiplexing → si fa una quality control per verificare se le reads sono venute bene → successivamente tramite alcuni software vengono rimossi gli adattatori in 5' e 3', però gli dico di conservare soltanto quelle molecole che hanno sequenziato tutto o in parte l'adattatore perché se gli dico di conservare tutte le molecole, anche quelle senza adattatore, è altamente probabile che conservi anche delle molecole derivanti da DNA contaminant → Merged FastQ (mettere insieme) sia delle reads Forward R1 e reads Reverse R2 fino a ricreare il filamento di origine di queste due reads → Clipped FastQ: cioè ho sequenze in cui sono state rimosse tutte le informazioni che non ci servivano → 2) READ MAPPING → le allineo ad un genoma di riferimento umano per vedere se ci sono delle sostituzioni nucleotidiche tramite dei software (in rosso) → posso rimuovere i duplicati di PCR: abbiamo visto che inizialmente facciamo target e archiment aumentando il numero delle reads che poi vado a sequenziare, quindi rimuovo i duplicati ottenendo un'unica sequenza Unique Seq read su cui fare le analisi → 3) Contamination check: le sequenze che devono essere analizzate hanno estensioni BAM e SAM e attraverso software (map damage ecc..) ho informazioni che riguardano i pattern di misincorporazione in 5' e 3' e una stima del DNA contaminante → 4) Genotyping: utilizzando software in cui ho la chiamata delle reads di consenso, cioè delle reads che hanno superato gli step della contamination check. Possono utilizzare eger per osservare gli SNPs presenti nelle sequenze.

Posso misurare anche la profondità della copertura, senza fare il genotyping e prima di fare il contamination check

Depth coverage: il coverage è il numero di volte che la regione di interesse è stata sequenziata, per cui mi indica quante volte quella posizione è coperta.

Se abbiamo a che fare con DNA mit che è circolare o DNA nucleare: ci sono dei software specifici



CircularMAPPER e MIA: vengono utilizzati per allineare genomi circolari, quindi DNAmIt.

*quelli in rosso vengono utilizzati per entrambi

ANGD e DICE: vengono utilizzati esclusivamente per il DNA nucleare

CONTAMIX e Schmutzi vengono utilizzati per il DNA mitocondriale

→ per la proporzione di reads autentiche.

FastQC -> controllo di qualità sui dati grezzi di sequenza. Fornisce un insieme di analisi da fare per avere una prima idea se i dati hanno problemi di cui si deve essere a conoscenza prima di fare qualsiasi ulteriore analisi successiva

ADAPER TRIMMING AND MERGING (RIMOZIONE DEGLI ADATTATORI E MERGING DI R1 E R2) → Permette di riunirle in un'unica reads.

R1 e R2 hanno lo stesso **READ IDENTIFIER** (stesso identificatore di lettura) → sono il reverse complement dello stesso frammento.

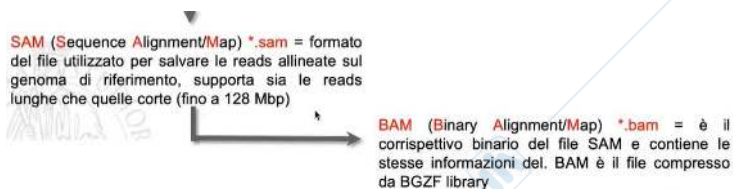
È importante una sovrapposizione delle reads merged di circa 30 paia di basi in questo modo si può ottenere una reads unica tra due differenti reads.

Adesso possiamo passare al **MAPPAGGIO o ALLINEAMENTO**

ALIGNMENT/MAPPING = determinare la posizione di ciascuna reads nel gen di riferimento

Mappare le reads su un genoma di riferimento è un compito molto complesso che dipende da svariati fattori, tra cui la variazione genetica nella popolazione, l'errore di sequenziamento, la lunghezza delle reads corte e la quantità di reads corte da mappare

Un software che viene utilizzato è **BWA** (Burros-Wheeler Aligner). Il file che otteniamo ha come formato **SAM** (Sequence Alignment/Map).



GENOMI CIRCOLARI

- BWA prova a mappare completamente le reads sequenziate sul genoma di riferimento e contrassegnare le reads che non possono essere mappate completamente come non mappanti
- Le ultime duecento basi di genomi circolari sono generalmente ricostruite male a causa dell'incapacità di mappare le reads in queste regioni

MIA (Mapping Iterative Assembler)
A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing

Richard E. Green¹, Anne-Sophie Malasomma², Johannes Krause³, Adrien W. Briggs⁴, Philip L.F. Johnson⁵, Gidon Katz⁶, Johannes Krause⁷, Lutz M. Meyer⁸, Thomas Martin⁹, John D. Stone¹⁰, Ray Stiller¹¹, Michael Schumacher¹², Benoit A. Briceau¹³, Michael Sorensen¹⁴, Jonathan M. Aitken¹⁵, Michael E. Brown¹⁶, Peter A. Underhill¹⁷, David Bradley¹⁸, Sarah Krause¹⁹, Paul Schumacher²⁰, Klaus Harigsen²¹, J. Daniel Wallau²², J. Daniel Wallau²³, J. Daniel Wallau²⁴, and Stephan Pääbo²⁵

CircularMapper

Genome Biology

SOFTWARE Open Access

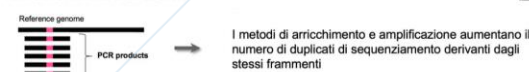
EAGER: efficient ancient genome reconstruction

Harshad Kulkarni^{1,2}, Cyrus Ghaffari^{1,2}, Alexander Heide^{1,2}, Alexander Sato¹, Christian Frey¹, Johannes Krause^{1,2}, and Ray Stiller^{1,2}

Metodo che consente il soft-clipping

- Tre differenti steps:
- 1) creare un genoma di riferimento "allungato"
 - 2) Mappaggio delle reads sul genoma allungato
 - 3) Rimappare le reads ambigue

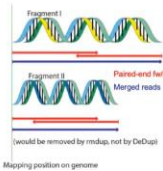
RIMOZIONE DEI DUPLICATI



Poiché la copertura di specifici loci genomici è importante per le analisi successive, i duplicati di PCR possono portare a falsi positivi a cause di un numero elevato di frammenti duplicati

Samtools rmdup

Viene presa in considerazione solo la posizione iniziale delle reads mappate



DeDup

Prende in considerazione entrambe le posizioni delle reads mappate (inizio e fine)



$$COVERAGE = \frac{\text{Throughput}}{\text{Region_Of_Interest}}$$

Throughput → #basi mappanti
 Region_of_Interest → Lunghezza delle regione d'interesse

Esempio: Studio sul mtDNA

Campione A = 500.000 bp mappanti	Campione A → Coverage = $5 \cdot 10^5 / 16569 = 30,176$
Campione B = 18.000 bp mappanti	Campione B → Coverage = $18 \cdot 10^3 / 16569 = 1,086$

Regione di interesse = 16569 bp

Attraverso il sequenziamento light

FastQC Clip&Merge → **BWA**

Preprocessing
 Adapter clipping
 Merging
 Low qual. filtering

Mapping
 Mapping of reads against Hg19

Accurate sex identification of ancient human remains using DNA shotgun sequencing
 Pontus Skoglund^{a,*}, Jan Storå^b, Anders Götherström^b, Mattias Jakobsson^{a,c}

$$R_Y = n_Y / (n_X + n_Y)$$

(n_Y) = number of alignments to the Y chromosome
 (n_X) = number of alignments to the X chromosome

R_Y ~ 0.0022 (Female icon)
R_Y ~ 0.09 (Male icon)

attraverso il mappaggio si può fare la determinazione del sesso.

Una volta che ho fatto i mappaggi devo in qualche modo sintetizzare tutte le informazioni in un file che sia utilizzabile, quindi devo concentrare tutta l'informazione in dei file che siano utilizzabili anche per le successive fasi di analisi di genetica di popolazione. Ci sono essenzialmente due formati:

VCF file: VARIANT CALL FORMAT è un formato che si può ottenere con vari software tra cui anche SAME TOOLS ecc. questo file è un file in cui vengono conservate soltanto le informazioni relative ai polimorfismi e alle varianti. Per cui avrò informazioni e avrò riportato in questo file soltanto quelle posizioni che sono variate rispetto alla reference: le posizioni identiche alla reference non vengono scritte. In questa prima parte del file (file in txt), ci sono le meta-

UNIVERSITÀ DEGLI STUDI FIRENZE

GENOTYPING

Le informazioni ottenute dal mappaggio devono essere riassunte in un formato di file utile

VCF
(Samtools, GATK, VcfTools)

Il Variant Call Format (VCF) è un formato specifico per salvare solo le varianti identificate. Il VCF è un formato file txt che contiene righe di meta-informazioni, una riga di intestazione e quindi una serie di righe di dati, ciascuna contenente informazioni su una specifica posizione nel genoma.

The screenshot shows a VCF file with two main sections: 'Meta-information lines' at the top, which include file paths and tool versions, and 'Data lines' at the bottom, which are tab-separated rows containing genomic coordinates, quality scores, and variant identifiers.

CONSENSUS SEQUENCE
(MIA, GATK, VcfTools/Bedtools)

Formato file Fasta che contiene tutti i nucleotidi identificati in ciascuna posizione in un allineamento di sequenza

informazioni, cioè le informazioni relative al campione che stiamo analizzando, mentre nella DATA LINES (sotto) ci sono le informazioni relative ai polimorfismi.

FASTA (comunemente trovato nelle banche genetiche): si utilizza quando invece si va a studiare genomi piccoli, per esempio quelli batterici o mitocondriali, e si va a ricreare la sequenza consenso. Quindi la sequenza consenso è conservata in un file FASTA (con il simbolino > all'inizio), e nella sequenza

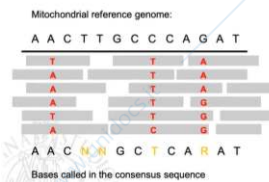
consenso sono riportate esattamente tutte le posizioni che ho ottenuto. per cui se vado a studiare il mitocondrio, nella sequenza consenso avrò 16.569 paia di basi sia che siano esse identiche alla reference, sia che siano variate. Quindi quello che conservo qua dentro è TUTTA l'informazione, mentre nel VCF file conservo solo l'informazione legata alle varianti.

RICOSTRUZIONE DELLA SEQUENZA CONSENSO

La si può fare sia su DNA mitocondriale, sia DNA nucleare.

DNA mitocondriale

-> Elevato numero di reads mappanti su ciascuna posizione



- Viene chiamata la base di riferimento se la qualità di tale chiamata è almeno 30 con una copertura minima della posizione è di almeno 5 reads
- Viene chiamata la variante se la qualità di tale chiamata è almeno 30 con una copertura minima di 5 reads e almeno il 90% delle reads presentano lo SNP

DNA mitocondriale

Quando vado a creare una sequenza consenso, che sia con l'uno o con l'altro tipo di file, come la creo? Come faccio a sapere che in quella posizione c'è esattamente una variante o c'è una reference? O c'è una base identica alla reference? Lo faccio andando a vedere come sono distribuiti i nucleotidi nel mio mappaggio: ci sono dei software che fanno le chiamate, non è che mi metto a vedere tutte le reads e a fare la chiamata. Utilizzo quindi dei software particolari: per il DNA antico, dove i parametri devono essere molto più stringenti, proprio perché ho molecole corte e danneggiate, utilizzo un aggiornamento del software GATC che è sempre fornito all'interno della piattaforma IGOR e lo imposto con questi particolari filtri.

Si imposta che: • Una chiamata venga effettuata SE e SOLO SE ho almeno 5 reads che mi coprono quella base, quindi al di sotto delle 5 reads non faccio la chiamata, ma metto una indeterminata, cioè una (N) = nomenclatura IUPAC che indica un sito in cui non ho chiamata.

• Se tutte e 5 (come minimo) le reads che ho portano un nucleotide come la reference, chiamo ovviamente la reference, significa cioè che in quella posizione non c'è variazione. In questo caso per esempio ho 4 reads che portano il nucleotide come la reference e 2 che invece portano la variante, quindi cosa chiamo? NON chiamo la variante perché per chiamarla deve essere presente almeno nel 90% delle reads presenti, quindi la variante la chiamerò quando ho 5 reads in un modo e solo una diversa.

• 50-50: metà portano un nucleotide e metà un altro, per esempio metà portano la A e metà portano la G: cosa chiamo in questo caso? Non si sa quindi in questo caso ci metto una (R) che nella nomenclatura IUPAC significa una indeterminazione tra due nucleotidi, in particolare indica una indeterminazione tra G e A, mentre per l'indeterminazione tra C e T si utilizza la (Y). Per cui quando vado ad aprire la mia sequenza consenso generata da GATC e vedo che in quella posizione c'è la R, vado ad aprire l'allineamento, ricerco la posizione ed effettivamente vedo che 50% delle reads portano una variante e 50% ne portano un'altra.

Alla fine la mia sequenza consenso viene ricostruita esattamente, utilizzando parametri che sono molto stringenti -> questo vale per il DNA antico perché ho l'esigenza di essere molto conservativa perché nella maggior parte dei casi quelle che potrebbero essere varianti non lo sono, ma sono misincorporazioni o lesioni post-mortem che la molecola ha subito, per cui devo essere molto conservativa e stringente nel ricreare la sequenza consenso.

DNA nucleare

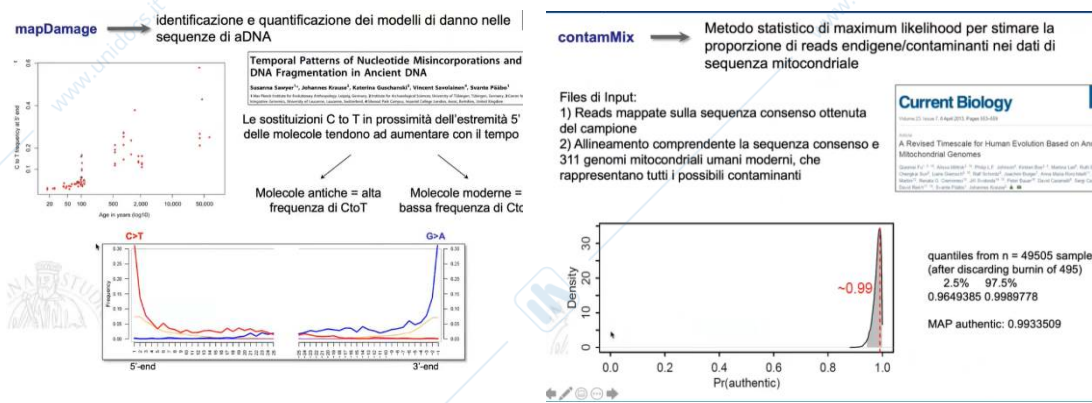
Si procede alla stessa maniera. Problema

Nuclear DNA

-> Numero molto basso di reads che mappano su ciascuna posizione

CONTAMINATION CHECK

Lunghezza media dei frammenti di DNA antico varia da 40 a 150 bp.



Il contaMIX ci serve per dire che probabilità c'è che quel determinato DNA mitocondriale che ho recuperato appartiene ad un'unica fonte biologica. Se appartiene ad un'unica fonte biologica → >95%.

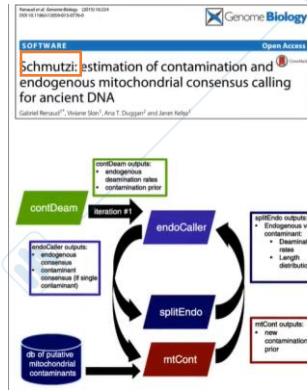
Il contaMIX può essere utilizzato sia per campioni degradati antichi che hanno problemi di contaminazione, ma anche per campioni degradati moderni.

Sequenza consenso antica:

- Frammenti corti
- Alta frequenza di C to T

• Proporzione di reads autentiche/contaminanti $\geq 94\%$

RISULTATO AUTENTICO



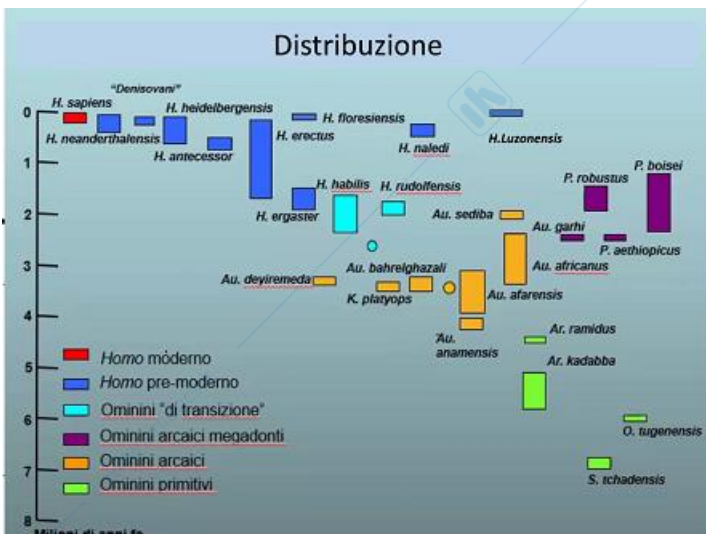
Approccio iterativo per stimare la contaminazione umana moderna in set di dati di DNA mitocondriale umano antico e contemporaneamente ricostruire il genoma mitocondriale endogeno

- Utilizzando i patterns di deaminazione delle reads e la distribuzioni della lunghezza dei frammenti, Schmutzi ricostruisce accuratamente la sequenza de genoma mitocondriale endogeno anche quando la contaminazione supera il 50%
- La contaminazione umana moderna viene stimata utilizzando un database di genomi mitocondriali umani mediante attraverso l'analisi delle posizioni diagnostiche

PMDtools utilizza i danni post-mortem (PMD), i punteggi di qualità delle basi e i polimorfismi biologici per identificare le sequenze di DNA degradate che è improbabile che derivino dalla contaminazione moderna

*utilizzati in genomi di carattere nucleare

LEZIONE 9.12



Questo schema è quello che conosciamo ad oggi delle linee che fanno parte della nostra storia evolutiva. Non su tutte si ha la possibilità di applicare le tecnologie di cui ha discusso fino ad ora. Poche delle specie riportati in questo schema hanno potuto essere studiati da un punto di vista molecolare, alcuni solo da un punto di vista morfologico.

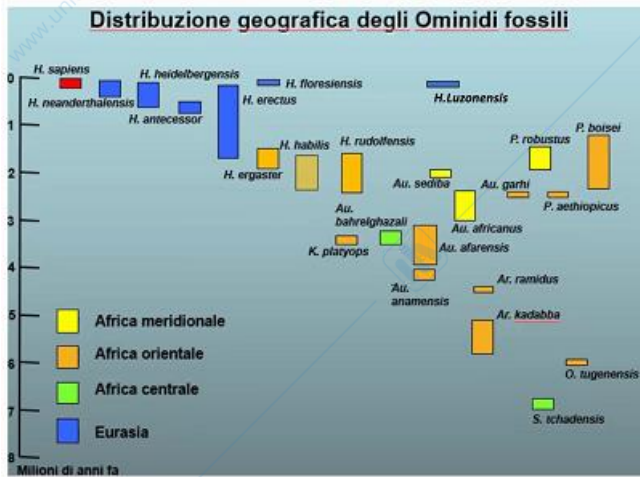
C'è una grossa variabilità, anche in termini di denominazione di specie.

Altra cosa da notare è la suddivisione che riguarda i differenti colori

Considerando una distribuzione di carattere orizzontale → contemporaneamente ci sono differenti specie anche nello stesso periodo temporale → molte specie che ora non ci sono più convivevano contemporaneamente.

Sono state fatte grandi scoperte grazie alle grandi tecnologiche che hanno riguardato le specie più recenti ottenendo informazioni interessanti sulla storia evolutiva della specie.

Grazie a queste tecnologie è stata messa in discussione la storia evolutiva della nostra specie sulla parte più recente.



Distribuzione geografica degli ominidi fossili fino a circa a 1 milioni 800 mila anni fa → i fossili sono stati collocati in Africa orientale, meridionale, centrale..

→ conferma che la linea che ha portato al genere Homo ad oggi si è concentrata sul continente africano.

Le prime tracce del genere Homo che si spingono fuori dall'Africa possono attribuirsi all'Homo Erectus..ecc.



Ritrovato in Ciad ca 6.8- 7 Milioni di anni fa e collocato in uno spazio geografico africano. È un ominide primitivo con caratteristiche ben delineate: sono forme di ominini che hanno un attività arboricola (vive sugli alberi), con un andatura bipede facoltativa → si iniziano a vedere le prime forme che hanno un andatura bipede. I ritrovamenti principali si hanno in Kenya, Ciad. I reperti scheletrici: arti superiori molto lunghi, arti inferiori che danno un'informazione sulla loro andatura bipede facoltativa.

Punti importanti :

Gli Ominini primitivi presentano alcuni caratteri morfologici più derivati che indicano postura eretta e andatura bipede e inizio riduzione denti canini, mentre altri caratteri sono primitivi.

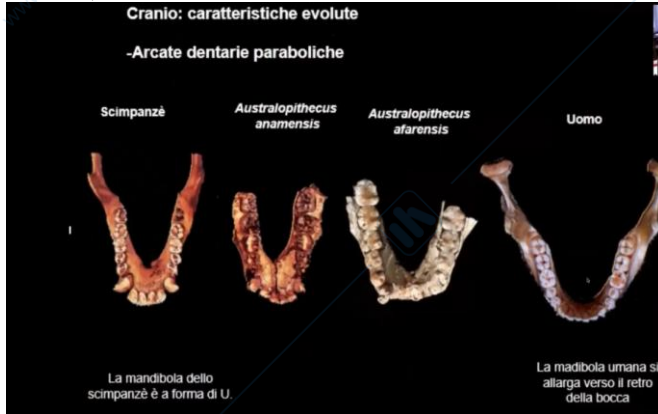
Quindi si possono considerare dei probabili antenati della linea evolutiva umana.

La postura eretta e andatura bipede sono da considerarsi un esempio di exaptations (o pre-adattamenti), cioè caratteri che hanno modificato la loro funzione originaria adattandola ad un altro uso.



Altro ritrovamento **Australopithecus anamensis**, 4.2-3,9 milioni di anni fa. È stato ritrovato anche il

cranio, con struttura molto ridotta.



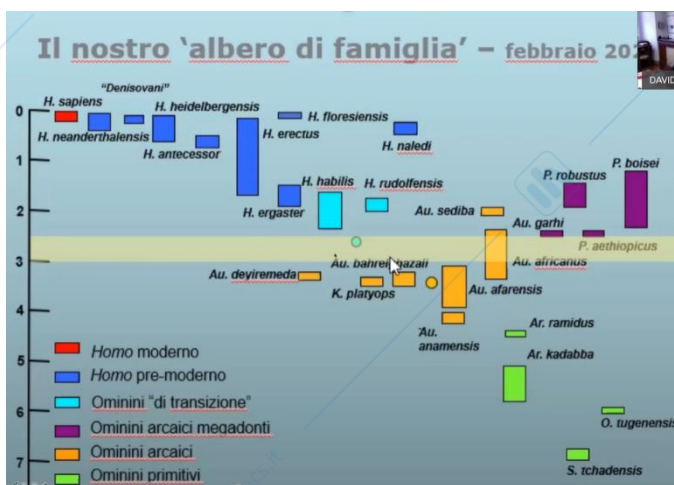
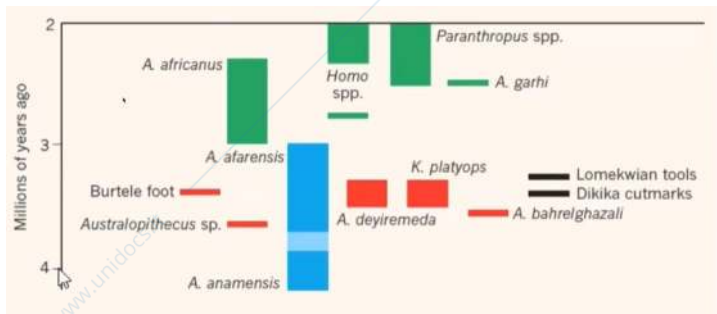
Cranio: molto piccolo

Mandibola: Cambia la forma della mandibola.

I canini dei scimpanzè sono molto sviluppati, mentre avvicinandoci all'uomo si ha una riduzione della struttura.

Nel corso di centinaia di migliaia di anni, tutta una serie di carattere paleoambientale, per sfuggire ai predatori e di recupero del cibo avrebbe portato un'andatura più bipede rispetto ad un'andatura arboricola, in termini di fitness.

Esisteva una certa variabilità di queste forme di **Australopithecus** e abbastanza lontane l'una dalle altre perché non è facile trovare forme fossili ben conservate in un ambiente africano con climi particolari.



stesso grafico di prima, ma con la striscia intorno a 3 milioni di anni fa.

→ I cambiamenti climatici sono responsabili della maggior parte di speciazioni ecc.. di forme che hanno vantaggi rispetto ad altri.
→ è ciò che è avvenuto intorno a 3 milioni di anni fa evidenziata dalla striscia temporale.

Infatti in quella striscia temporale si possono trovare degli ominini di transizione la cui attribuzione è al genere Homo → inizia a comparire il genere Homo → Questa

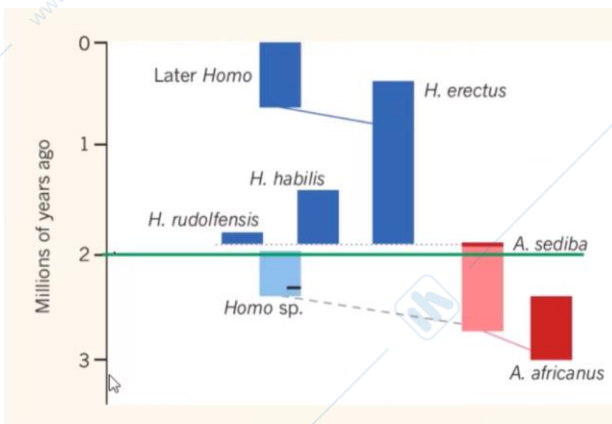
comparsa può essere dovuto a vantaggi di carattere evolutivo dovuto a cambiamenti climatici molto

importanti.



Scoperta di una nuova specie: **Australopithecus sediba** → indica che le forme di **Australopithecus** all'interno del continente africano erano abbastanza variabili → forse dovuto alla loro particolare approccio a differenti nicchie di carattere ecologico.

Queste forme molto adattate ad un particolare tipo di ambiente → se l'ambiente cambia improvvisamente → li porta ad estinzione.



In seguito al cambiamento climatico cominciando a comparire le prime specie afferenti al genere Homo e iniziano a scomparire le forme di **Australopithecus**.

Non solo cambiamenti climatici, ma anche fattori astronomici: cambiamento dell'inclinazione rispetto all'asse terrestre, cambiamento della direzione ecc.. → tutti una serie di fattori che hanno portato a cambiamenti climatici importanti che hanno dato un vantaggio a specie che erano più adatte a questi cambiamenti climatici



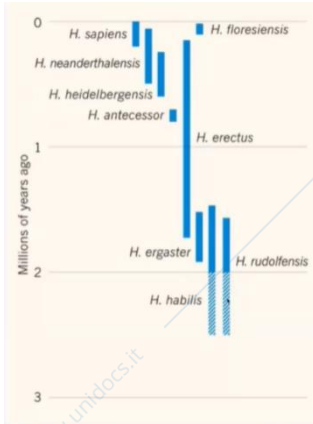
→ Portando all'affermazione di una sola specie del genere Homo: Sapiens.

OMININI ARCAICI MEGADONTI: Le cosiddette Australopithecine Robuste

Vissute tra 2,5 e 1 milione di anni fa: sono caratterizzate da denti molto più grandi → alimentazione di radici molto dure. La caratteristica peculiare del Paranthropus robustus è una cresta sagittale dove si attaccano muscoli masticatori

Intorno a 2.5- 3 milioni di anni fa cominciando a comparire le prime specie del genere Homo.

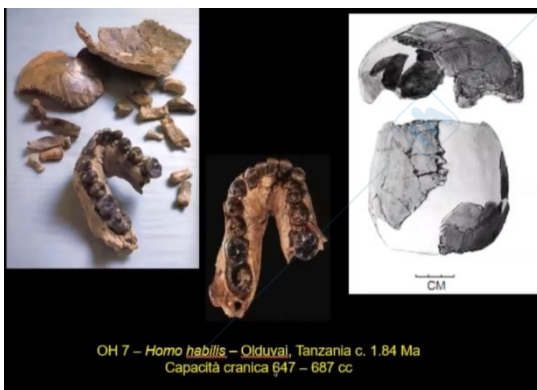
La prima specie del genere Homo è **H. Habilis**, chiamato così perché in concomitanza di queste forme ci sono anche delle forme di carattere culturale: strumenti costruiti per poter attingere a delle risorse. Uno dei ritrovamenti è la mascella, molto ampia datata intorno a 2,3 milioni di anni fa in Etiopia →



In realtà le prime forme del genere Homo non era Habilis, ma non è stato determinante la specie → grazie a ritrovamenti in Etiopia intorno a 3 milioni di anni fa di un mascellare differente più simile al genere Homo.

Ci sono più forme inerenti al genere Homo: **H. rudolfensis** 2.5 3 milioni di anni fa in Malawi → con un arcata del mascellare sempre più aperta, ascrivibile al genere Homo.

Olduvai George e Koobi Fora in Africa dove ci sono stati molti ritrovamenti relativi al genere Homo Habilis



Capacità cranica è il doppio rispetto ai precedenti.

Un aspetto interessante di Homo Habilis è la parte postcraniale → sia gli arti superiori e inferiori hanno caratteri primitivi (tibia e fibula) → sono molto più simili nella parte corticale agli individui che avevano ancora la capacità di arrampicarsi sugli alberi → quindi alcuni autori la considerano ancora **Australopithecine**

Antichi utensili attribuibili al genere Homo (2 milioni e 700 mila anni) utilizzati per scarnificare gli animali. È molto importante perché il fatto di accedere a risorse animali → la possibilità di poter accedere a queste risorse ci permette di avere una possibilità di assumere un cibo più nutriente →

assumere un cibo più nutriente permette di avere più risorse per un cervello che forse si sta ingrandendo, uno degli organi a cui serve più energia.

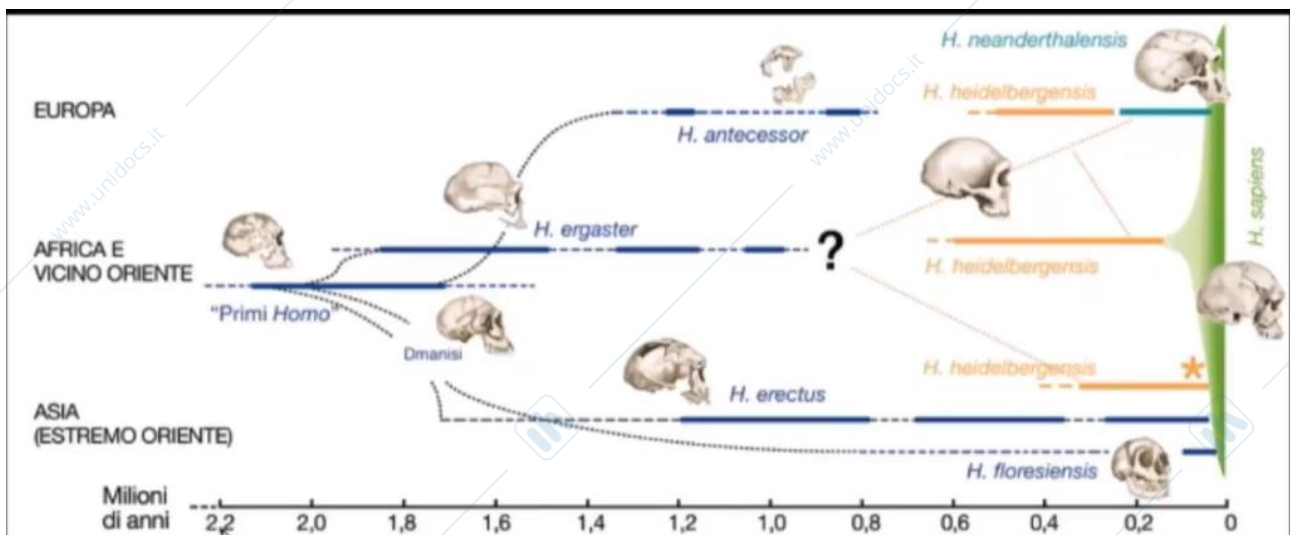


Con il genere Homo si assiste ad una prima diffusione al di fuori dell'Africa, circa 1,8 milioni di anni fa: da parte di forme di Homo più evolute, come Homo Ergaster, o Erectus: con un'andatura bipede non più facoltativa e una struttura imponente che permetteva di fare percorsi molto lunghi.

Ritrovamento in Georgia 1,77 milioni di anni fa una forma battezzata come georgus → forma appartenente al genere Homo: è una scoperta importante perché ci indica una prima diffusione del genere Homo che è arrivata in Georgia circa 1,7 milioni di anni fa da parte di Homo Georgus.

continuare

LEZIONE 14.12



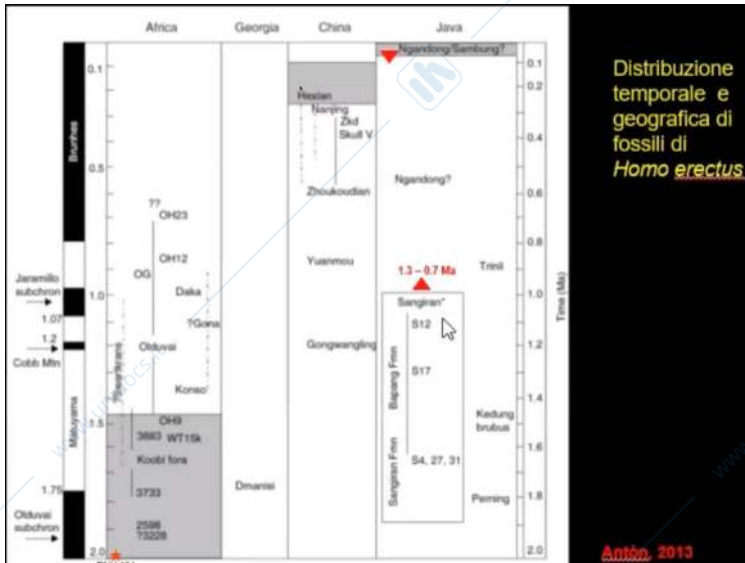
Come si può descrivere come tentativo l'uscita del genere Homo dall'Africa circa intorno a 1 milione e 700 mila anni fa.

La prima forma di Homo che esce dall'Africa è **H. Ergaster** grazie alle loro caratteristiche fisiche che può prendere differenti nomi in base al luogo geografico in cui ci troviamo.

Un primo reperto ritrovato in Georgia è H. georgus intorno a 1,7 milioni di anni fa → quindi questa fu la prima uscita dall'Africa su una rotta orientale che vede H. Georgus ritrovata in Georgia. Da queste

forme nelle varie ondate di fuoriuscite intorno a questo periodo, per quanto concerne alle forme asiatiche facciamo riferimento all'H. Erectus (se mi dite che in africa c'era H. erectus anziché ergaster va bene lo stesso).

Ora ci concentriamo sulla parte asiatica:



DISTRIBUZIONE TEMPORALE E GEOGRAFICA DI FOSSILI DI HOMO ERECTUS. Considerando quest'altro schema: si può verificare come queste forme si trovano sia in Africa (quindi può essere considerato come H. ergaster), in Cina, in Georgia e nell'isola di Java.

Le frecce rosse indicano una nuova datazione : 1,3 0,7 milioni di anni fa Ma della gran quantità di resti fossili riferibili a Homo di Ngandong, Homo erectus di Java

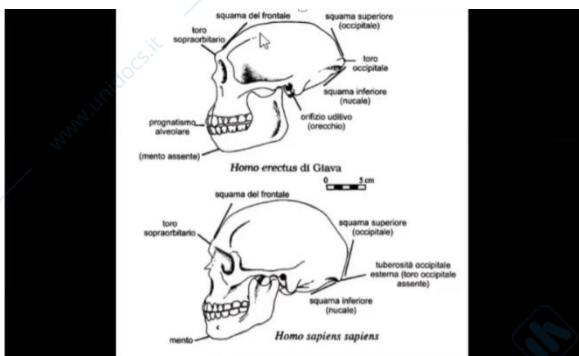
Tuttavia i pochi resti ritrovati fanno

pensare che non sia stata un ondata migratoria importante come invece lo è stata quella della specie Sapiens, che è la TERZA OUT OF AFRICA. Si può notare un aumento della capacità cranica da Homo Habilis, passando per H. rudolfensis a Homo. Ergaster.

Quindi **Homo ergaster**: è la forma di Homo forse più vecchia, 1 milione e 700.000 anni fa in Africa ed è la specie che per prima ha deciso di lasciare l'Africa.

CARATTERISTICHE HOMO ERGASTER: Postura eretta con arti inferiori più lunghi rispetto agli arti superiori → indica la possibilità di percorrere percorsi su due gambe e anche possibilità. Ha un cranio allungato una regione soprorbitale molto importante che si osserva in tutte le forme arcaiche del genere Homo e diminuzione del prognatismo,

assenza del mento in Homo Erectus.

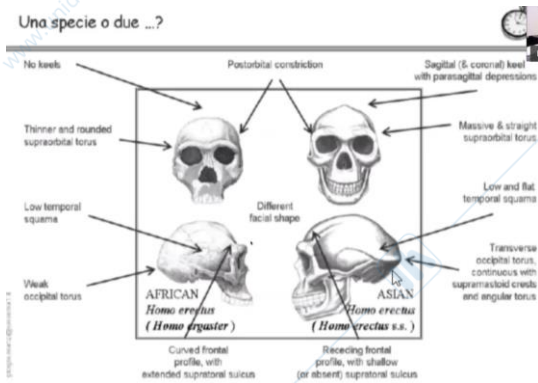


Crani di *Homo erectus* di Giava e di uomo attuale robusto (*H. sapiens sapiens*). Tra le differenze più importanti di *H. erectus* si notano il toro supraorbitario prominente, il frontale più basso e sfuggente, l'occipitale più angolato e l'accentuato prognatismo alveolare.

HOMO ERECTUS O ERGASTER una specie o due specie?

Considerando le forme africane si può considerare una forma di continuità della forma ergaster nella forma erectus

Se consideriamo Erectus in Asia, fuori dall'Africa si può considerare come una specie differente dalle caratteristiche morfologiche



Un aspetto importante dell'Homo erectus è l'aspetto culturale: cambiare modo di costruire gli oggetti con utensili acheuleani bifacciali caratterizzata da manufatti litici a forma di mandorla e lavorati su due lati in modo simmetrico → il discorso della simmetria è importante perché denota la capacità di poter costruire oggetti con una certa precisione → quindi era capace di costruire strumenti più precisi → SVILUPPO DELLE POTENZIALI CEREBRALI. Infatti con Homo erectus si ha un aumento delle dimensioni cerebrali. (1200 centimetri cubi).

Quindi in Homo erectus si osserva:

- **modifica della struttura corporea**
- **aumento delle dimensioni corporee**
- **aumento delle dimensioni del cervello**
- **utilizzo di utensili in pietra più sofisticati**
- **dieta di più elevata qualità → carne, midollo, fegato, pesce** grazie ad evidenze di ossa rotte di grandi erbivori e mangiato il midollo → porta ad un aumento proteico → è importante perché mangiare proteine significa una qualità di dieta elevata.
- **Ipotesi degli acidi grassi polinsaturi essenziali (es. Omega 3) che si trova nei pesci**

Il cervello più grande ha bisogno di più energia e quindi si può una dieta diversa rispetto a prima.

C'è la possibilità, documentato 1 milioni di anni fa in Africa di utilizzo non occasionale, ma sistematico del fuoco → perché la cottura permette una migliore digeribilità del cibo → significa che l'apparato masticatore si modifica ulteriormente diminuendo lo spazio e quindi diventando più semplici e piccoli → permette una maggiore espansione delle dimensioni del cervello

Quindi l'aumento della capacità cranica nell'ERECTUS era dovuto a questi due aspetti: in incremento della dieta più variegata e una diminuzione dei muscoli masticatori.

Ci sono delle evidenze della fuoriuscita del genere Homo erectus dall'Africa anche in ITALIA a PIRRO NORD (in Foggia) 1.3 milioni anni fa e SIMA dell'elefante (SPAGNA) 1,2 milioni di anni fa. → era il primo reperto scheletrico attribuito al genere Homo con un'arcata mascellare aperta attribuibile al genere Homo.

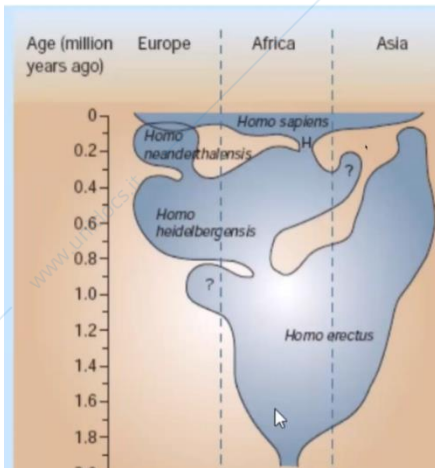
PIRRO nord è un sito datato ad 1,6- 1,3 milioni anni fa in cui si osservano industrie di tipo 1.

1,2 milioni anni fa si hanno diverse tracce di popolamento in Europa del genere Homo.

HOMO antecessor: forma ominide che si è rilevata essere afferente al genere Homo che lascia l'Africa dopo H. Erectus e arriva in Europa intorno a 1,2 milioni di anni fa e da origine a Homo antecessor.

Quindi abbiamo detto che la fuoriuscita del genere *Homo erectus* che nella prima Out of Africa va a colonizzare l'Asia e alcune forme si sono viste anche in Europa.

Forse è esistita anche una seconda migrazione con una forma più evoluta intorno a 1,2 milioni di anni fa. FORME DI *HOMO HEIDELBERGENSIS* dovute all'evoluzione di *Homo ergaster* all'interno dell'Africa 700 mila-800 mila anni fa avrebbe portato ad una seconda espansione da parte di questa forma che avrebbe lasciato l'Africa e colonizzato l'Europa dando successivamente luogo a *Homo antecessor* e successivamente a *Homo Neanderthalian*.



Ultima fuoriuscita dall'Africa intorno a 60-70 mila anni fa molto importante da parte di *Homo Sapiens*, considerata un'evoluzione successiva africana del *HOMO HEIDELBERGENSIS* che colonizza Europa e Asia andando a mescolarsi con altre forme che prima erano al di fuori dell'Africa: Neanderthaliani o *Erectus*.

È una prima idea di come si possa essere sviluppata l'evoluzione recente che descrive i movimenti migratori in base alle ricostruzioni fossili, ma grosse informazioni si sono avute con l'utilizzo delle molecole di DNA.

HOMO HEIDELBERGENSIS: è stato ritrovato in Germania ca 600 mila anni fa.

Conosciamo il genoma mitocondriale e nucleare

CLADOGENESI VS ANAGENESI: due modi per interpretare l'evoluzione della nostra specie del genere *Homo*.

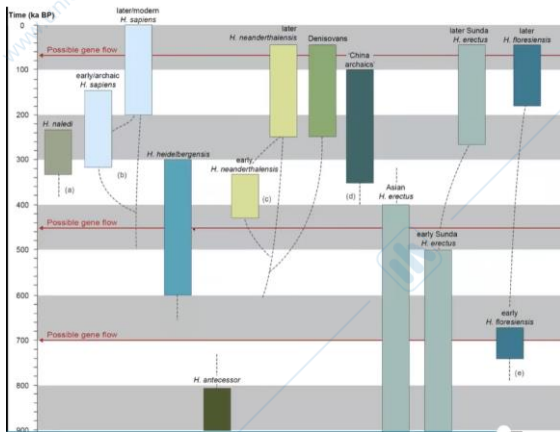
CLADOGENESI: è un meccanismo di carattere evolutivo dove si hanno colli di bottiglia dove poi man mano si sviluppano popolazioni differenti, a cladi

ANAGENESI: è una continua evoluzione da una specie all'altra.

Inoltre ci possono essere state nel nostro continente grazie a dati di carattere paleoclimatico eventi di estinzione e ripopolamento da parte del genere *Homo heidelbergensis* numerosi volte e possono aver determinato cambiamenti di carattere popolazionistico, di carattere genetico e in una successiva ripopolazione hanno potuto portare a cambiamenti che hanno portato successivamente dell'*Homo* di Neandertal intorno ai 350 mila anni fa in Europa.

In Europa fra 600 mila e 400 mila anni fa → popolamento in Europa si fa importante grazie alla seconda fuoriuscita (2 out of Africa) dell'*H. Heidelbergensis*. è M

Una delle scoperte importanti → In Spagna sono state ritrovate in Sima de los Heusos ritrovamenti di resti fossili → 17 crani appartenenti probabilmente a 17 individui differenti in una grotta con strumenti particolari di ca 430 mila anni fa auspicabili a *Homo heidelbergensis*



HOMO NALEDI: 335-236 mila anni fa, ritrovata in sud Africa. È afferente al genere Homo ma con caratteristiche particolari: Volume del cranio molto piccolo essendo anche una specie molto recente → ha caratteristiche del genere Homo e Australopithecine

DENISOVA: è una forma rinvenuta in una grotta dei Monti altai di 50-30 mila nni fa. È stato ritrovato un dente, molto robusto (non attribuibile a Neandertal o erectus) e un piccolo frammento dell'osso della mano → analizzando il genoma mitocondriale si è scoperto che non era ne Neandertaliano, ne sapiens → si trattava di una specie nuova, ma che ad oggi non ne conosciamo le caratteristiche post craniali e craniali, ma si conosce il genoma.

È stata ritrovata successivamente una mandibola 160 mila nni fa su delle alture → appartenente ai denisoviani → probabilmente una popolazione adattata a altezze importanti.

Sull'isola di Flores (indonesia) è stato rinvenuto una nuova specie HOMO **floresiensis** ribattezzata HOBBIT per le sue dimensioni molto piccole, afferenti al genere HOMO. Non è stato possibile estrarre il DNA.

Sono state ritrovate molti individui di questa specie con dimensioni molto piccoli (altezza di 1 metro ca) → probabilmente si tratta di endemismo insulare → un adattamento a determinati contesti di carattere ambientale → perché priva di predatori, carnivori.

Alcuni di questi reperti sono stati datati 18 mila-20 mila anni fa abbastanza recenti e alcuni molto antichi di 800 mila anni fa.

HOMO LUZONENSIS: Rinvenuto a Callao Cave nelle isole filippine filippine, datazione 50 mila nni fa.

Quindi c'era una grossa variabilità del genere Homo, che oggi non esiste più, è scomparsa.

NEANDERTAL: forme estinte fossili afferenti al nostro genere. Sono le più recenti e ci sono un numero elevato di reperti fossili.

Il primo reperto è stato ritrovato nel 1856 nella valle del fiume NEANDER.

Si conosce bene l'anatomia, sviluppo, genetica, stile di vita, comportamento, ma non si conosce il motivo dell'estinzione. Si trovavano in Europa principalmente, un po' in medio oriente, sui monti altai. **NON SONO AFRICANI!!** Non ci sono mai stati!!!

CARATTERISTICHE NEANDERTAL:

- toro sopraorbitario
- ciglion occipitale

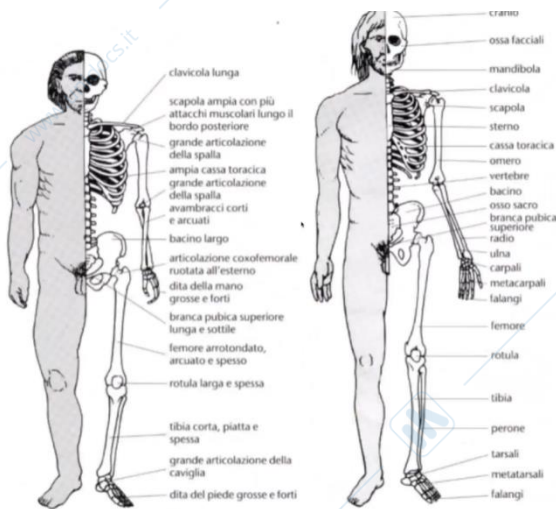
- forma allungata del cranio
- assenza del mento

Lo si riconosce benissimo dall'Homo sapiens in quanto ha il mento, non ha il toro supraorbitario, cranio globulare.

“Modello di accretion” model di origine dei neandertaliani: dall'Homo heidelbergensis → spiegazione di carattere morfologico → dovuto a estinzioni e parziali e ripopolamenti dovuti a cambiamenti climatici molto importanti.

- Capacità cranica media 1172 a 1740 centimetri cubi → un grosso cervello ma con un'espansione dei lobi parietali e antero-posteriore

In sapiens si ha un'espansione verticale dei lobi parietali e globularizzazione a fronte di un'espansione laterale dei globi frontali.



A livello anatomico i Neandertal hanno una robustezza e particolari caratteristiche dello scheletro a basse temperature

LEZIONE 16 DICEMBRE

Uomo di altaura, video.

ALIMENTAZIONE

I neandertal hanno un grosso cervello, i sapiens hanno vissuto con i neandertal in Europa.

Attraverso lo studio del dental calculus (tartaro) presenti su denti Neandertaliani abbiamo capito di cosa si potessero cibare e scoperto anche il comportamento di alcuni di questi individui.

È possibile capire quali sono stati gli alimenti di cui i Neandertaliani facevano uso estraendo il DNA dal dental calculus di vari Neandertaliani, solo da due si ottengono informazioni.

Neandertal più continentali avevano una dieta più carnivora

Neandertal di El Sidron si nutriva prevalentemente di alimenti di foresta (pinoli, funghi).

Altro aspetto importante: è stato trovato all'interno del dental calculus di questo Neandertaliano il DNA di pioppo → il pioppo contiene acido acetilsalicylico → probabilmente serviva come antiinfiammatorio → infatti questo neandertaliano presentava un ascesso → quindi ha fatto capire che i neandertaliano conosceva alcune piante medicinali → elevato psichismo.

ESTINZIONE: perché non ci sono più? Tante teorie

- Competizione con H. Sapiens di carattere ecologico
- Cambiamenti climatici che av

rebbero diminuito la fitness dell'organismo in quanto aveva un sistema corporeo adatto a climi freddi, ma sappiamo che sono vissuti anche in periodi caldi

- Malattie → portate dai Sapiens

JABEL IRHOUD: è stato ritrovato un cranio di 315 mila anni fa in AFRICA → molto simile a heidbengensis → fa pensare che l'ultima fuoriuscita di Homo sapiens dall'Africa sia avvenuta 60-70 mila anni fa. L'evoluzione dei Sapiens pare che si sia sviluppata da queste forme di individui heidberngensis africane, anche se sono state documentate altre fuoriuscite prime dei 60-70 mila anni fa, grazie ai reperti ritrovati fuori dall'Africa

Mentre forme più antiche di heidbengensis avrebbe dato vita a forme di Neandertal fuori dall'AFRICA

Mentre la diffusione (diffusione perché pochi reperti) del genere homo Sapiens intorno a 160 mila anni fa.

Quindi si può dividere la diffusione di Homo Sapiens: Early dispersals (120-60 mila) e later dispersals (<60-30 mila). Ci possono essere eventuali forme di mescolamento tra Neand e Sapiens e con Denisova.

Se nella linea temporale aggiungo il DNA più antico che con la metodologia classica siamo riusciti a tirare fuori, mi torna che abbia 175.000: dell'uomo di Altamura (neandertal) e quindi con la metodologia classica possiamo studiare: SAPIENS, NEANDERTAL, HEIDELBERGENSIS, HOMO DI FLORES e forse qualche individuo di Homo erectus, ma non le altre perché sono molto più antiche e quindi con la metodologia classica non siamo capaci di poterle studiare. Ad oggi però con la metodologia classica è stato possibile analizzare solo 2 specie che sono: sapiens e neandertal.

Ci sono 4 modelli con cui si può spiegare la HUMAN EVOLUTION:

1. **MODELLO DEL CANDELABRO**: Questo modello dice che quando l'homo ergaster uscì per la prima volta dall'Africa, intorno a 1 milione e 500 mila anni fa, andò in Asia, in Europa, in Austral-Asia e lì, generazione dopo generazione, si è evoluto in forme umane differenti tra di loro, fino a portare alle forme umane attuali: asiatici, australiani, europei ecc. e in questo percorso di carattere evolutivo, tutte le forme umane che vengono trovate nei diversi continenti non sono altro che una evoluzione delle forme precedenti. Esempio: l'homo di neandertal, secondo il modello del candelabro, sarebbe l'antenato delle popolazioni europee. Modello che regge pochissimo perché questo presupporrebbe che la variabilità genetica osservata in ciascun continente fosse talmente elevata da distinguere un individuo africano da uno asiatico o da uno australiano, cosa che sappiamo non è possibile, in quanto il genere Homo all'interno dell'Africa è molto recente

2. **MODELLO MULTIREGIONALE** (dell'evoluzione multiregionale) Ci dice le stesse cose del modello precedente, ma ci dice anche molte cose in più: include i flussi genici tra i vari continenti. Questo, dal punto di vista degli scambi genici tra popolazioni, è plausibile perché continui scambi genici portano ad una omogeneizzazione dei vari popoli e quindi sarebbero spiegate anche le differenze che non si osservano oggi tra popolazioni diverse, però si prevede che i neandertal siano i nostri diretti antenati.

3. **MODELLO DEL REPLACEMENT** Dice che è vero che la specie appartenente al genere Homo, in particolare l'ergaster, è uscito dall'Africa 1 milione e 700.000 anni fa, ma è anche vero che c'è stata una seconda ondata massiccia di sapiens fuori dall'Africa, intorno ai 70.000 anni fa, con cui, nei vari continenti, si sono sostituite le varie comunità che prima ci vivevano. Quindi quando sono arrivati i sapiens, l'homo di neandertal si sarebbe estinto, senza dare al sapiens nessun contributo genetico.

4. **MODELLO DELL'ASSIMILAZIONE** Dice la stessa cosa del modello del REPLACEMENT, ma dice anche che in alcuni casi è possibile che ci sia stata una sorta di contributo genetico delle popolazioni che prima erano fuori dall'Africa con le popolazioni che sono uscite fuori dall'Africa

Con l'analisi del DNA mitocondriale il modello che funzionava di più era il modello del replacement perché il DNA mitocondriale neandertaliano e Sapiens erano differenti.

MA..con l'era dell'NGS → ci ha aperto nuovi orizzonti → è stato analizzato DNA di individui vissuti 400 mila anni fa e abbiamo analizzato DNA nucleare di Sapiens, Neandertal e heidelbergensis e il DNA di una nuova specie Denisova → quindi nel pleistocene superiore 50 mila anni fa esistevano diverse specie → sequenziamento completo dell'homo di neandertal → ci siamo accorti che hanno con noi condiviso 1-4% con noi. Guardando il DNA delle pop africane non condividono DNA con i neandertal, quindi sicuramente ci siamo incrociati con loro dopo la fuoriuscita dall'afrika e non hanno quindi un'origine comune.

379 soggetti europei e 286 dell'estremo oriente

L'apporto genetico dei Neanderthal all'uomo moderno sarebbe avvenuto in due differenti momenti: il primo subito dopo la migrazione degli uomini moderni fuori dall'Africa, che avrebbe interessato l'intera popolazione uscita da quel continente, e il secondo, successivo, che avrebbe interessato solo le popolazioni migrate verso l'estremo oriente; nell'uomo moderno si è complessivamente conservato - sia pure sparso su individui diversi - il 20 per cento del genoma neandertaliano

frequenza elevata di geni neandertaliani nelle regioni cromosomiche che influenzano le caratteristiche della pelle e dei capelli, geni che potrebbero aver contribuito all'adattamento all'ambiente freddo non africano.

alleli ascrivibili ai Neanderthal e correlati al rischio di malattie (in particolare lupus, cirrosi biliare, malattia di Crohn, maculopatie, diabete di tipo due), e un gene noto per essere correlato a una maggiore dipendenza dal fumo.

povere di genoma neandertaliano sono le regioni cromosomiche che contengono geni espressi nei tessuti testicolari

il cromosoma X nel suo complesso

pur essendo una specie strettamente imparentata, alcuni dei geni dei Neanderthal evidentemente non sono stati "tollerati", provocando una ridotta fertilità degli ibridi maschi

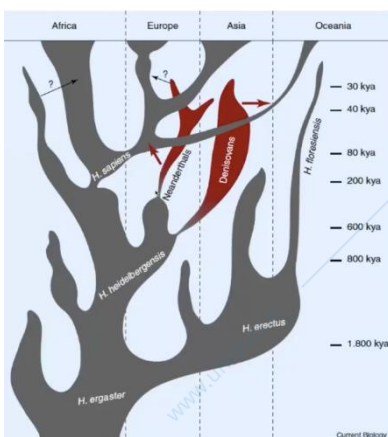
perché quando abbiamo analizzato il DNA mit dei neandertaliani e confrontato con i sapiens sono differenti → sappiamo che si eredita per via materna → molto probabilmente non tutti gli ibridi avevano lo stesso grado di fertilità: molto probabilmente gli ibridi che avevano una fertilità maggiore erano gli ibridi che avevano maschio neandertal e femmina sapiens → la femmina che nasceva da questo ibrido aveva alta probabilità di avere la prole fertile rispetto al maschio → questo si spiega perché esistono pochi geni legati al cromosoma Y dei neandertaliani in noi Sapiens.

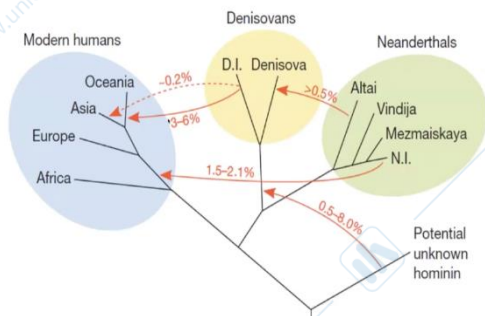
Ma in qualche combinazione particolare c'era una certa fertilità perché non sarebbe stato possibile osservare geni neandertaliani in noi → qui viene meno il concetto di specie nel genere Homo in queste ultime fasi dell'evoluzione.

Sappiamo che anche i denisova si sono incrociati sia con sapiens che con neandertal. I denisoviani erano una specie differente grazie all'analisi del DNA mitocondriale differente da Neandertal e Sapiens.

h. heidelbergensis esce fuori dall'afrika intorno ai 700mila anni fa → 2 out of africa

Molto probabilmente da origine ai denisoviani in asia, ai neandertal in Europa e Homo sapiens in africa, che poi esce dall'afrika e incontra i Neandertal con i quali si incrociano in periodi diversi. Prosegue migrazione sapiens in asia che si incrociano con denisoviani, questo si osserva bene nelle pop della nuova guinea dove si trova fino al 6% del genoma denisoviano.



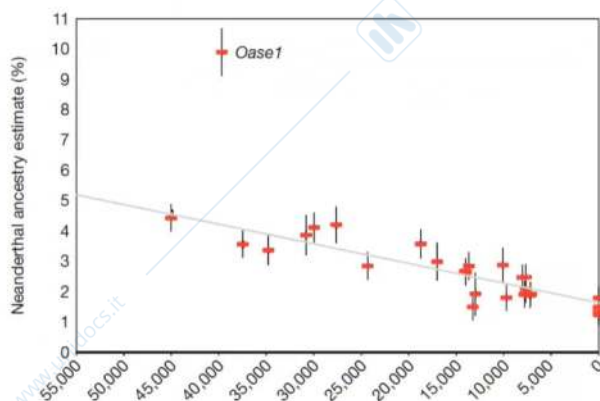


è interessante che una popolazione non conosciuta è esistita ed ha fornito da 5 all'8% del suo genoma alle popolazioni denisoviane. Quindi il modello ad oggi più corrispondente è quello dell'assimilazione.

Anche lo studio del mitocondrio degli scavi a Sima de los Huesos, cioè *Homo heidelbergensis*, aspetto interessante che hanno il DNA mitocondriale più simile ai denisoviani che ai neandertali. Com'è possibile? Se abbiamo detto che è una specie che ha dato origine alle popolazioni neandertali??

Quando abbiamo analizzato il DNA nucleare di Sima de los Huesos c'ha detto che il DNA è più simile ai neandertali. Allora perché il mtDNA dell'*Homo* di Sima de los Huesos è più simile ai denisoviani perché molto probabilmente il mtDNA con cui abbiamo confrontato fino a oggi i neandertali, non è il vero DNA mitocondriale che avevano 40.000 anni fa, può essere stato diluito da incroci con *Sapiens* che erano usciti dall'Africa. L'ha dimostrato Cosimo Posth con il mtDNA di un neandertale che è vissuto prima dei recenti neandertali ha DNA mitocondriale differente rispetto agli ultimi neandertali. Quindi quello più recente ha influenze di *Sapiens* che è arrivato dall'Africa e ha diluito il mitocondrio neandertalico rendendolo differente dai veri neandertali quindi vissuti intorno ai 100.000 anni fa.

La genetic history of Ice (storia genetica degli uomini dell'età del ghiaccio): se si osserva quanto neandertal c'è in *Sapiens*, osserviamo che i geni neandertalici si osservano che vanno scomparendo → quindi c'è una selezione negativa che viene via via eliminata.



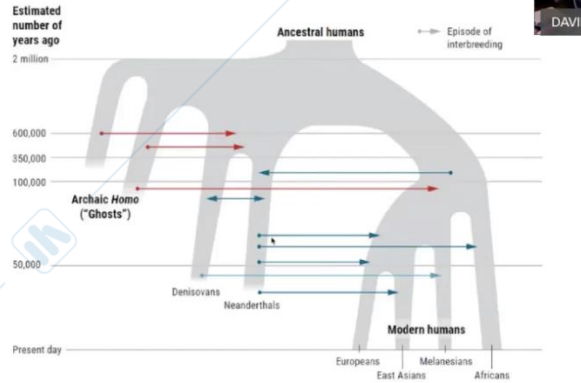
Non c'è DNA neandertalico in Africa, solo alcune, probabilmente per retroespansione.

Analizzato un individuo con metà genoma neandertalico e metà *Sapiens*, F1. → ha dimostrato che ci sono stati incroci.

Le **analisi statistiche**, basate su complessi modelli di ereditarietà, tassi di mutazione e incrocio di popolazioni, hanno restituito **8 possibili scenari** capaci di spiegare l'assortimento genetico osservato nelle popolazioni odierne. La spiegazione che i ricercatori hanno ritenuto più plausibile è quella che prevede un **antico evento di ibridazione** tra gli antenati di Neanderthal e Denisova e una popolazione "super arcaica" eurasiatica, probabilmente rappresentata da *erectus*, avvenuta fuori dall'Africa tra i 700.000 e i 600.000 anni fa (prima freccia rossa in alto a sinistra in figura).

Sappiamo infatti che negli esseri umani di oggi è presente il Dna di **almeno altre due popolazioni** di cui non sappiamo granché, e che per questo chiamiamo **popolazioni fantasma** (*ghost population* - potrebbe trattarsi di *erectus* o qualche suo stretto parente ancora non identificato). Anche con loro, a un certo punto, ci siamo accoppiati e c'è stato uno **scambio orizzontale di geni** (freccia rossa lunga in figura).

Sappiamo che Denisova e Neanderthal si sono incrociati tra loro perché abbiamo avuto la fortuna di trovare i resti fossili di **Denisva**, una **ragazza** di almeno 13 anni, vissuta circa **90.000 anni fa** nella caverna di Denisova sui monti Altai in Siberia, **figlia di un genitore Neanderthal e uno Denisova** (fraccia blu con doppia estremità in figura).



GRAPHIC: VERINOT ET AL./SCIENCE. ADAPTED BY V. ALTOUNIAN/SCIENCE. DATA: A. ROGERS AND S. SANJARJAMAN

Altro muro che le nuove scoperte genetiche hanno abbattuto è la convinzione, saldamente presente fino a pochi mesi fa, che solo gli esseri umani moderni che vivono **fuori dall'Africa** fossero portatori di Dna neanderthaliano. È stato invece scoperto, a febbraio di quest'anno, che anche **popolazioni africane odierne** possiedono una piccola porzione di **Dna neanderthaliano** (freccia blu da neanderthal a popolazioni africane). Ciò dovrebbe significare che alcune popolazioni di **sapiens euroasiatici** portatrici di Dna neanderthaliano sono tornate in Africa e hanno

Noi **sapiens**, come si è detto, ci siamo ibridati sia con Neanderthal sia con Denisova. Noi portiamo loro tracce nel nostro **genoma**, ma anche loro probabilmente ne portavano di nostre nel loro **Dna mitocondriale**, segno di un incrocio avvenuto più di 100.000 anni fa. Alcuni ricercatori fanno risalire questo evento di **introgressione** fino a **270.000 anni fa** (freccia blu da destra a sinistra).

LEGGI MEGLIO

CONSIDERAZIONE CONCLUSIVE di H. SAPIENS:

Considerazioni conclusive

- Diffusione in tutte le zone abitabili della terra
- Episodi ripetuti e differenziati in direzioni diverse, di entità diversa e con impatti diversi
- Su questi informazioni derivanti dalla genetica di popolazioni e archeologia
- Impatto ecologico di queste diffusioni
- Cambiamento tipo di economia
- Economia di caccia e raccolta
- Economia di produzione - agricoltura e allevamento - transizione Neolitica
- Cambiamento nello stile di vita