



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN "SCIENZE BIOMEDICHE, ENDOCRINOLOGICHE E NEUROFISIOLOGICHE"

CICLO XXII

COORDINATORE Prof. Martini Alessandro

**Ricerca di nuovi polimorfismi allelici per l'identificazione personale e
caratterizzazione di diverse emoglobine animali ed umane per la
discriminazione di specie.**

Settore Scientifico Disciplinare BIO/13

Dottorando

Dott. Faccinetto Christian

(firma)

Tutore

Prof. Tognon Mauro

(firma)

Correlatore

Dott. Pizzamiglio Marco

(firma)

INDICE

Introduzione.....pag. 3

Scopo.....pag. 22

Materiali e Metodi.....pag. 25

Risultati.....pag. 47

Discussione..... pag. 71

Ringraziamenti.....pag. 74

Bibliografia.....pag. 75

INTRODUZIONE

Le discipline Biologiche e Biomediche, come la Biologia Cellulare e la Genetica Molecolare, hanno assunto un ruolo sempre maggiore per facilitare l'identificazione personale impiegando le diverse conoscenze che si sono man mano acquisite nel tempo, come ad esempio nell'ambito del Progetto Genoma Umano. Queste conoscenze e le competenze tecniche derivate dall'impiego delle varianti alleliche polimorfiche hanno consentito di introdurre anche nell'ambito delle Scienze Forensi uno strumento indispensabile per la risoluzione di casi giudiziari.

Il progredire delle conoscenze scientifiche nei settori della Biologia, Genetica e Biotecnologie hanno reso possibile l'analisi della variabilità umana a livelli sempre più accurati e precisi; nel tempo si è così passati dallo studio della diversità a livello biologico, ovvero ricerca delle variabilità dei gruppi sanguigni come A,B,0, Rh ecc., degli antigeni leucocitari HLA, delle sieroproteine, enzimi, ecc. all'analisi molecolare di sequenze altamente variabili del genoma, dette polimorfismi del DNA. In pratica si è passati dallo studio delle variabilità fenotipiche, cioè le caratteristiche osservabili di un individuo, come le caratteristiche facciali, la forma delle orecchie, l'analisi delle impronte digitali, alla variabilità genotipica direttamente dove essa risiede, ovvero nella molecola di acido deossiribonucleico: il DNA.

I caratteri ereditari, dedotti sia dall'analisi del genotipo che del fenotipo, ci permettono di caratterizzare un'identificazione personale. Per questo scopo le analisi devono essere indirizzate a marcatori con specifiche caratteristiche come:

- elevata variabilità per poter essere differenziati, vale a dire essere altamente polimorfici e quindi consentire un profilo individuale certo;
- avere un modello di ereditarietà conosciuto, come ad esempio Autosomico Dominante o Recessivo, legato al cromosoma X oppure legato ai geni mitocondriali (eredità materna);
- essere altamente conservati nel genoma nucleare e mitocondriale dell'individuo, così come nelle generazioni successive;
- essere in unica copia all'interno del genoma.

L'identificazione personale nell'uomo è il risultato della comparazione di dati sia Fenotipici che Genotipici ottenuti dall'analisi di campioni in esame con quelli di controllo. Nel caso in cui esista una compatibilità totale tra il campione in esame ed il controllo, l'analista può confermare che con elevata probabilità possa trattarsi della stessa persona. Viceversa, se esiste discrepanza tra i campioni in esame ed il controllo potremmo dire che non si tratti dello stesso individuo. In base alle frequenze alleliche di singoli marcatori genetici calcolate nella popolazione la compatibilità viene espressa come una probabilità. Le possibilità che esista un'altra persona con lo stesso profilo genetico diminuisce proporzionalmente aumentano il numero dei marcatori polimorfici analizzati. In tal modo, i risultati ottenuti con questi studi sul DNA permettono di identificare un individuo con una probabilità molto vicina al 100%.

Lo studio dei polimorfismi del DNA per l'identificazione personale ebbe i suoi esordi all'inizio degli anni '80. In effetti questi studi e le analisi del DNA hanno rivoluzionato le conoscenze sul modo di identificare e discriminare tra loro gli individui. Infatti, analizzando un numero sufficiente di regioni di DNA polimorfiche, si può osservare una grande variabilità tra i profili genotipici degli individui, riducendo al contempo il grado di omologia a livelli sempre più bassi. La tipizzazione del DNA quindi permette di diminuire grandemente la probabilità di avere due profili identici in due diversi individui. Questa probabilità può ridursi a dei livelli talmente bassi che la si converte in un metodo di identificazione e non solo in un metodo di inclusione/esclusione.

La rivoluzione scaturita dall'analisi dei polimorfismi ipervariabili del DNA si accompagna ad uno studio sempre più analitico del fenotipo come ad esempio la caratterizzazione del sistema multiallelico del gruppo sanguigno A,B,0, così come di altri gruppi sanguigni e delle emoglobine corrispondenti.

La scoperta del polimorfismo dei gruppi sanguigni umani A,B,0 (Landsteiner K., 1900) ha favorito l'applicazione del concetto di variabilità genetica nella nostra specie. Sulla membrana cellulare dei globuli rossi di ogni individuo sono presenti diversi antigeni, ossia macromolecole altamente immunogeniche in grado di evocare una risposta anticorpale. La reazione tra antigene ed anticorpo produce uno specifico immunocomplesso che può essere a sua volta identificato. Nella figura 1 vengono rappresentati i possibili fenotipi dei diversi eritrociti di gruppo A, B, AB e 0. I rispettivi antigeni A, B,

AB e 0, sono in grado di reagire con sieri immuni che riconoscono questi specifici marcatori favorendo quindi la formazione degli immunocomplessi. L'eccezione è data dagli eritrociti di gruppo 0 privi di antigeni. In questo caso non si formeranno immunocomplessi. Questa scoperta che risale all'inizio del secolo scorso, è esemplificativa di un sistema genetico multiallelico molto semplice. Tuttavia, il sistema A,B,0 è utile ad esempio per escludere un individuo dalla scena de crimine, mentre risulta non adeguato per verificarne la presenza. Tale aspetto risulta più cogente nel caso del gruppo 0, presente in circa il 40% della popolazione.

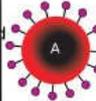
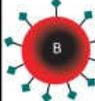
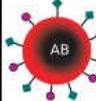
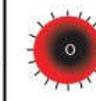
	Group A	Group B	Group AB	Group O
Red blood cell type				
Antibodies present	 Anti-B	 Anti-A	None	 Anti-A and Anti-B
Antigens present	A antigen	B antigen	A and B antigens	No antigens

Figura 1: Gruppi sanguini

Restando nel nostro esempio di un indagato di crimine, le indagini sono tese a stabilire se il sospettato era effettivamente presente sulla scena del crimine. L'analisi del genotipo furono rese possibili a partire dagli anni '80, quando si acquisirono nuove conoscenze sul DNA e nuovi metodi di studio (come ad esempio l'elettroforesi su gel di agarosio o di poliacrilammide per fare uno studio dei profili del genoma). Tali analisi hanno permesso di individuare diversi polimorfismi del genotipo. Tuttavia, in una prima fase di studio ed applicazione il numero ridotto e il tipo di marcatori genotipici impiegati non consentivano una alta variabilità e di conseguenza non erano particolarmente informativi. I dati ottenuti, denominati con il termine anglosassone "Match Probability" cioè la probabilità che due persone non imparentate condividessero lo stesso profilo genetico, risultavano troppo alti, compresi in un intervallo tra 0,01 e 0,001.

Nel caso di alleli ereditati indipendentemente, la Match Probability si calcola moltiplicando le specifiche frequenze alleliche presenti nel genoma in esame. Va da sé che maggiore è il numero di alleli presi in esame, maggiore l'eterozigosità ad ogni gene/locus, più basso sarà come risultato il valore di Match Probability.

Tuttavia ci sono situazioni in cui il Match Probability può aumentare (Jobling and Gill, 2004) come nel caso in cui:

- il profilo risulta parziale a causa della degradazione del campione di DNA con la conseguente riduzione di loci informativi disponibili;
- un sospetto condivide per parentela molti alleli con il colpevole; ad esempio come nel caso di due fratelli che condividono il 50 % del genotipo;
- un sospetto e il colpevole appartengono alla stessa popolazione costituita da così pochi individui tale per cui è probabile che queste persone siano consapevolmente o inconsapevolmente imparentati tra loro. In questo caso oltre ad avere molti polimorfismi genotipici in comune avranno anche uno stesso profilo del DNA mitocondriale per eredità materna o del cromosoma Y (in medicina forense questi aspetti vengono denominati relativi ad una sottopopolazione).

Nel 1985 vi fu un autentico punto di partenza con l'introduzione da parte di Alec Jeffreys, del metodo denominato "DNA fingerprinting". Per questo scopo Jeffreys impiegò sequenze di DNA ipervariabili, chiamati minisatelliti (Jeffreys A.J., 1985). La scoperta era relativa ad alcune regioni di DNA con sequenze che si ripetevano in tandem, una dopo l'altra, più volte. Il numero delle ripetizioni presenti differisce tra individuo e individuo. In seguito sono state sviluppate tecniche in grado di discriminare le variazioni del numero di queste sequenze ripetute di DNA in base al peso molecolare del frammento di DNA che le conteneva. In questo modo Jeffreys creò i presupposti per la nascita del test di identificazione personale attualmente ancora in uso ed in seguito sviluppato da altri autori che hanno aumentato il numero delle sequenze ipervariabili.

Le sequenze ipervariabili o minisatelliti o Variable Number Tandem Repeats (VNTR), sono delle sequenze nucleotidiche della lunghezza di 9-100 paia di basi (bp) ripetute in tandem un numero variabile di volte, centinaia o anche migliaia, e in più loci cromosomici (Jeffreys A.J., 1985, Figura 2).

Studi eseguiti sul genoma umano hanno dimostrato che le VNTR mappano preferenzialmente in regioni preterminali dei cromosomi umani.

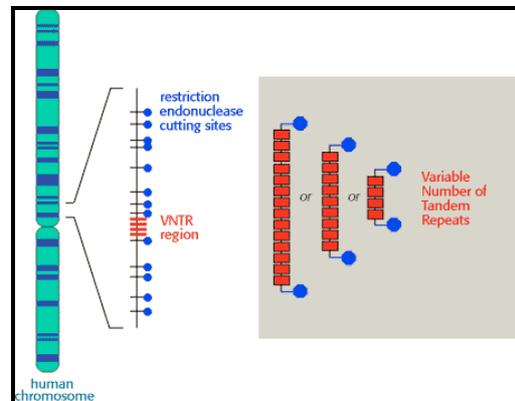


Figura 2: VNTRs

Nell'analisi del genoma umano dove vengono esaminati i VNTR, veniva principalmente impiegata la metodica denominata Polimorfismi di Lunghezza dei Frammenti di Restrizione o RFLP (Figura 3). Tale analisi impiegava uno o più enzimi di restrizione che tagliano il DNA in corrispondenza di sequenze palindromiche presenti in prossimità delle VNTR o minisatelliti.

Questa analisi RFLP prevede inizialmente la digestione del DNA genomico con uno o più enzimi di restrizione seguita da elettroforesi in gel di agarosio per separare i frammenti di DNA in base al loro peso molecolare (PM), seguita o meno da ibridazione molecolare su filtro chiamata tecnica di Southern blot. In questo caso i frammenti di DNA separati su gel vengono trasferiti su una membrana di nitrocellulosa o nylon. L'analisi procede impiegando in ibridazione su filtro una sonda fluorescente o radioattiva specifica per le sequenze ripetute in tandem o VNTR o DNA minisatellite, che si appaierà per complementarità dei nucleotidi alla sequenza di DNA omologo. Come ultima fase dell'analisi l'ibrido molecolare DNA – DNA verrà svelato mediante autoradiografia. Una volta sviluppato il film l'immagine che ne scaturisce mostrerà il profilo sia elettroforetico che di ibridazione, corrispondente al genotipo dell'individuo.

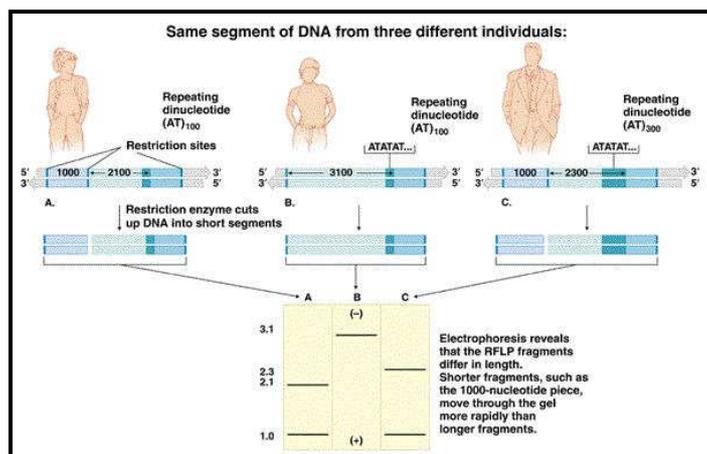


Figura 3: Tecnica VNTR-RFLP

L'impiego di sonde molecolari costituite da segmenti di DNA complementari alle sequenze ipervariabili o VNTR o minisatelliti consente di produrre profili di ibridazione del DNA che in autoradiografia si palesa con la produzione di specifiche bande a diverso peso molecolare. Tale profilo di ibridazione permette quindi di identificare in modo praticamente univoco un individuo. Per questo motivo, vista l'unicità del profilo di ibridazione di un dato genotipo, le bande corrispondenti sono state chiamate "impronte digitali" del DNA. Con una singola sonda (single-locus probe) la Match Probability risulta minore di 3×10^{-11} , mentre usando due sonde (multi-locus probe) questo valore diventa minore di 5×10^{-19} . Perciò, con un multi-locus DNA fingerprinting, utilizzando quindi più di una sonda, ogni profilo di ibridazione di DNA, escludendo gemelli omozigoti, è proprio di un singolo individuo. Su base statistica non ci sarà alcun altro individuo nell'ambito dell'intera popolazione mondiale con il suo stesso profilo. Considerato che ogni individuo è caratterizzato da un numero specifico di ripetizioni nei minisatelliti per ogni locus considerato, la variabilità totale nel numero di ripetizioni risulta molto alta all'interno della popolazione mondiale. Il quadro complessivo che scaturisce dall'analisi del DNA quindi, è relativa ad un'impronta genotipica specifica del soggetto. I minisatelliti possono essere considerati come le impronte digitali. Infatti, così come è praticamente impossibile che due persone abbiano le stesse impronte digitali, allo stesso modo ogni individuo, avendo i propri minisatelliti, possiede una sua impronta genetica o del DNA. In questo contesto è bene ricordare che l'impronta genetica è molto più

analitica rispetto all'impronta digitale. Infatti, i minisatelliti sono presenti nel DNA in coppia, una copia ereditata dal padre e l'altra dalla madre. Questo consente, non solo di verificare che l'impronta genetica è unica in quanto derivata sia dal DNA paterno che dal DNA materno, ma anche di stabilire chi sono i genitori, i fratelli o i nonni facendo un confronto fra i vari genotipi.

I minisatelliti hanno trovato quindi applicazioni nelle analisi di paternità, e di conseguenza tale impronta genetica viene utilizzata anche nelle prove giudiziarie, se del caso. Inoltre ci preme ricordare come da una macchia di sangue o da un frammento di pelle rinvenibili sul luogo del delitto, è possibile estrarre il DNA, analizzarlo come descritto sopra e definire con certezza la sua provenienza, comparandolo con quello di una persona indiziata o della stessa vittima ed eventuali altri controlli. La tecnica del DNA fingerprinting (Gill P. 1985) fu utilizzata per la prima volta per risolvere il caso di due episodi di violenza sessuale e omicidio di due ragazze nel 1983 e 1986. In questo esempio la polizia locale constatò che le caratteristiche delle aggressioni erano simili e si cominciò a sospettare che vi fosse un unico colpevole che aveva agito a distanza di tre anni. Il presunto colpevole fu individuato in base a diverse prove, come ad esempio: 1. la presenza dello stesso profilo genetico trovato sulla scena dei due omicidi avvenuti a tre anni di distanza l'uno dall'altro; 2. venne escluso un sospetto, risultato poi un mitomane, che confessò gli omicidi in quanto il suo DNA non coincideva con quello del presunto assassino. Questo dato ci ricorda che in questo esempio ci fu il primo successo dell'uso dell'analisi del DNA nel dimostrare l'innocenza di un indagato. 3. Il Forensic Science Service organizzò il primo screening di massa su 4.000 soggetti, di paesi limitrofi per il test del DNA. L'indiziato dei crimini, pur cercando di evitare questo screening, fu identificato e scoperto. Infatti, l'analisi del suo DNA mostrò un profilo identico al DNA rinvenuto sul luogo degli stupri e degli omicidi consentendo in maniera inequivocabile di riconoscerlo come il colpevole dei delitti sopra detti.

La tecnica del DNA fingerprinting è molto efficace, ha un elevato potere discriminante, cioè ha un'elevata capacità di distinguere le differenze tra individui, soprattutto se si impiegano gli RFLP "multi-locus probe". Tuttavia queste analisi hanno un tempo di applicazione piuttosto elevato compreso fra le 6/8 settimane. Un altro limite è dato dalla esigua quantità di DNA o dal DNA degradato disponibile. Infatti, la tecnica del DNA fingerprinting non risulta più attuabile in assenza di una congrua

quantità di DNA nativo (circa 1 microgrammo minimo). L'avvento di tecnologie più avanzate ha favorito lo sviluppo di metodi basati sulla metodica denominata "polymerase chain reaction" (PCR). Questa tecnica consente di semplificare ed abbreviare l'analisi del DNA favorendo l'amplificazione di uno specifico frammento di genoma che verrà in seguito analizzato come una frazione del DNA. In poche parole non dobbiamo più prendere tutto il genoma, digerirlo con gli enzimi di restrizione, farlo correre con l'elettroforesi, trasferirlo su filtro, ibridarlo con una sonda, per verificare la presenza delle specifiche sequenze da identificare. Quindi l'utilizzo dei marcatori può avvenire in poche ore ottenendo risultati comparabili dal punto di vista della capacità discriminante.

In precedenza abbiamo descritto l'impiego dei minisatelliti o VNTR. Più recentemente sono stati identificati ed impiegati dei nuovi marcatori denominati microsatelliti o short tandem repeats (STR). Analogamente ai minisatelliti, i microsatelliti sono costituiti da sequenze ripetute di DNA non codificante, formate da unità di ripetizione molto corte di 1-6 bp (Figura 4). Questi STR si trovano dispersi in tutto il genoma, senza distinzione tra regioni codificanti e non-codificanti e all'interno dei geni senza distinzione tra esoni e introni. Gli STR mappano in prossimità e in loci associati frequentemente all'estremità 3' delle sequenze Alu (sequenze palindromiche presenti nel genoma umano riconosciute dall'enzima di restrizione Alu I). La loro posizione è probabilmente dovuta alla degenerazione della sequenza di terminazione dei geni (Economou et al. 1990, Beckman e Weber 1992). I loci STR presentano un alto grado di polimorfismo di lunghezza dovuto alla variazione nel numero delle unità ripetute (Litt e Luty 1989, Weber e May 1989). Oltre alla variazione nel numero delle ripetizioni, le sequenze STR possono variare anche nella lunghezza delle unità ripetute.

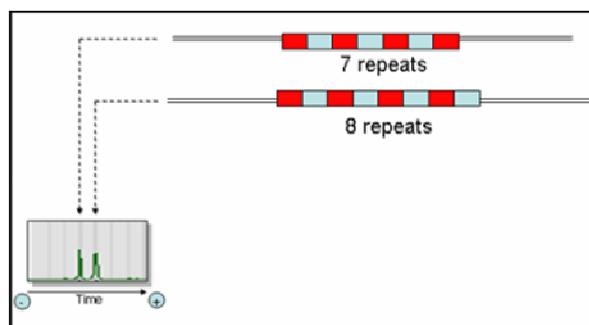


Figura 4: STRs

I microsatelliti o STR si classificano in base al numero di nucleotidi presenti nella ripetizione. Gli esempi sono riferibili alla presenza di 1-6 nucleotidi nella ripetizione. Quelli formati da mononucleotidi (poliA/poliT) sono associati alle sequenze Alu, alle long interspersed elements (LINE) e a sequenze simili ai retrotrasposoni e retrovirus, definiti anche retro elementi. Questi specifici STR non sono molto utilizzati a causa delle piccole differenze tra la grandezza degli alleli. Gli STR dinucleotidi rappresentano lo 0.5% del genoma, sono usati nelle analisi anche se hanno tassi di mutazione superiori agli altri. Gli STR trinucleotidi e tetranucleotidi (molto utilizzati perché possono essere facilmente discriminati), penta nucleotidi e esanucleotidi non sono molto usati a causa della loro scarsa presenza nel genoma umano. Un altro sistema distinzione degli STR, oltre alla discriminazione per la lunghezza, è data dal tipo di microsatellite:

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

Le conoscenze attuali indicano che i loci STR sono distribuiti in due gruppi principali: quelli che mostrano un piccolo numero di alleli ben differenziati (<12) e quelli che ne mostrano un numero elevato (>35) ma la cui differenza tra i vari alleli è più complessa. Gli STR di questo secondo gruppo sono molto più polimorfici; questi loci sembrano essere molto più complessi e hanno alleli multipli che si differenziano per il numero di ripetizioni, presentando una notevole eterogeneità. Queste caratteristiche non escludono a loci complessi di essere inclusi come marcatori per la identificazione genetica, ma differenze tra gli alleli in relazione alla variazione di una sola base, richiedono un'indagine molto analitica. Le frequenze di mutazione (inserzione/delezione) dei microsatelliti sono state stimate grazie all'analisi diretta verificandola sul campo (Brinkmann et al., 1998, Xu et al., 2000), vale a dire misurando direttamente le mutazioni sul DNA ottenuto da un'ampia casistica di popolazione umana. Per questo scopo gli autori hanno esaminato i discendenti di molte generazioni, studi di mutazione meiotica (Heyer et al., 1997). Le caratteristiche dei processi di mutazione sono state studiate anche in pazienti

con cancro al colon che mostravano instabilità del DNA a livello dei microsatelliti (Di Rienzo et al., 1998) e anche per comparazione tra i microsatelliti umani e quelli dello scimpanzè (Webster et al., 2002). Quello che si è ottenuto sono frequenze di mutazione (inserzione/delezione) di 10^{-3} / 10^{-4} per locus per generazione. Da questi studi si determina che:

- la maggior parte delle mutazioni (>85%; Brinkmann et al., 1998, Xu et al., 2000) implicano un aumento o una diminuzione di una ripetizione completa STR;
- la frequenza totale delle mutazioni aumenta quando aumenta o quando diminuisce troppo la lunghezza del STR. Comunque, il tipo di mutazione, così come la frequenza della stessa, dipende dalla lunghezza: sotto un determinato numero di ripetizioni la frequenza di mutazione diventa molto bassa. Questo spiega perché la lunghezza dei microsatelliti ha una distribuzione stabile e anche perché microsatelliti molto lunghi (>50 ripetizioni) sono molto rari;
- dinucleotidi mutano molto più velocemente rispetto ai tri/tetra nucleotidi (Chakraborty et al., 1995; Webster et al., 2002).

I loci STR possono essere amplificati usando la PCR e primers complementari alle regioni di sequenza unica adiacenti alle unità ripetute. L'elettroforesi in gel di poliacrilammide (Sambrook et al., 1989) usata in combinazione con la colorazione con argento (Bassam et al., 1991) permette l'analisi precisa dei frammenti STR amplificati, anche se attualmente esistono metodi ancora più precisi, affidabili e rapidi per il loro studio.

L'assegnazione della grandezza degli alleli (peso molecolare) viene attuata anche grazie al confronto con dei ladders allelici che consistono in una combinazione di sequenze amplificate di vari alleli che si usano come marcatori per ogni locus STR. La comparazione dei campioni con il ladder allelico (di cui sono noti il peso molecolare e la nomenclatura degli alleli presenti) permette una determinazione visuale rapida dei valori di grandezza, senza la necessità di fare calcoli addizionali.

In genetica molecolare quando si sceglie un STR per la costruzione di kit di analisi nell'applicazione forense, si richiede che questo abbia specifiche caratteristiche, come:

- alta eterozigosità;
- basso tasso di mutazione;
- devono dare origine alla minore presenza possibile di artefatti durante l'amplificazione;
- facilmente amplificabili anche insieme ad altri STR;
- di piccola grandezza (100-500 pbs).

L'uso dei microsatelliti presenta grandi vantaggi rispetto ad altri sistemi: la grandezza dei prodotti amplificati è piccola (100-500 pb) ed è perciò possibile ottenere risultati a partire da DNA degradato (frammenti anche minori di 1.000 pb). La piccola grandezza delle unità amplificate permette lo studio simultaneo di vari loci mediante amplificazione con PCR-multiplex; la tecnica della PCR permette di ottenere anche milioni di copie di un frammento di DNA in poche ore, consentendo così l'analisi di campioni la cui quantità è scarsa. Un altro vantaggio è, quindi, il tempo breve necessario per completare l'analisi grazie all'uso della tecnica di PCR, tempo che si è ridotto ancora di più con l'uso della tecnologia automatizzata basata sulla fluorescenza. Combinando la tecnica della PCR multiplex e l'analisi automatica con il tipizzatore con l'impiego di fluorocromi si ottengono i profili di DNA ancora più rapidamente.

La nomenclatura utilizzata per identificare i diversi loci STR è abbastanza esplicita. Originariamente, se un marcatore faceva parte di un gene o cadeva all'interno di un gene, era il nome del gene stesso che veniva usato per dare il nome al microsatellite. Per esempio il marcatore microsatellite TH01 è posizionato all'interno del gene Tirosine Hydroxylase che mappa sul cromosoma 11. La sigla "01" nel nome TH01 deriva dalla regione, contenente il numero variabile di ripetizioni, situata all'interno dell'introne 1 del gene tirosina idrossilasi. A volte il prefisso HUM è incluso nel nome del locus per indicare che fa parte del genoma umano, quindi solitamente il marcatore TH01 è indicato come HUMTH01.

Tutti i marcatori collocati in regioni extra-geniche o meno, oggi vengono definiti con la loro posizione sul cromosoma. Ad esempio il microsatellite D5S818 rappresenta un marcatore che non si trova

all'interno di un gene. In questo caso la lettera D sta per DNA; il successivo numero 5 si riferisce al cromosoma 5. DYS19 e DXS6800, invece appartengono rispettivamente al cromosoma Y e X. La lettera S indica che il marcatore di DNA è a singola copia, mentre il numero finale indica l'ordine secondo il quale i marcatori sono stati scoperti e catalogati su un particolare cromosoma.

I cromosomi X e Y, fatte le debite eccezioni, determinano rispettivamente il sesso femminile e maschile. Il cromosoma Y è il più piccolo tra i cromosomi umani, con una grandezza di 60 Mb (Morton, 1991), ed è formato da:

- regioni di eterocromatina: braccio lungo o Yq in posizione distale. E' composto da sequenze altamente ripetitive, con polimorfismi di lunghezza;
- regioni di eucromatina: braccio corto o Yp, nel centromero e nella zona vicino al braccio lungo.

La sua grandezza è abbastanza costante (≈ 30 Mb); qui si ha la presenza di geni, regioni di omologia con il cromosoma X e con gli autosomi e sequenze ripetitive tipiche del cromosoma Y.

Nel cromosoma Y si possono quindi differenziare 3 regioni di interesse: due pseudogeni (PAR-1 e PAR-2) e una regione specifica del cromosoma Y (Figura 5). Durante la meiosi, in Profase I durante il crossing-over, i cromosomi X e Y possono appaiarsi e ricombinare. Questo fenomeno avviene a livello di piccole regioni che contengono sequenze omologhe presenti su entrambi i cromosomi (regioni pseudoautosomiche). Delle regioni pseudoautosomiche, PAR-1 è la maggiore del cromosoma Y, con una grandezza di ≈ 2.6 Mb. Questa regione ricombina sempre, perché una ricombinazione a livello di questa regione è necessaria per una segregazione normale dell'X e dell'Y in meiosi (Ellis e Goodfellow, 1989a). PAR-2, la regione pseudoautosomica minore, è grande ≈ 0.32 Mb e non sempre partecipa alla ricombinazione; in più, la sua attività di ricombinazione non può andare a sostituire quella della PAR-1 (Freije et al., 1992) non essendo perciò né necessaria né sufficiente per l'esito corretto della meiosi maschile.

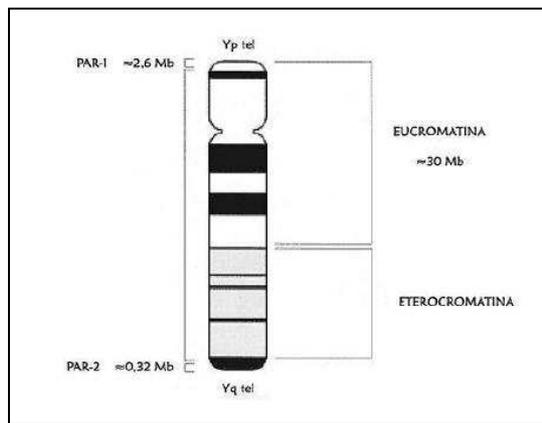


Figura 5: Cromosoma Y

Non sempre è possibile l'analisi dei loci microsatelliti (STR). Infatti quando si dispone di campioni antichi o altamente degradati, i risultati ottenuti dal DNA nucleare possono essere non attendibili. In queste situazioni si ricorre spesso al DNA mitocondriale, che è presente in numero elevato di copie per cellula e quindi ha una maggiore probabilità di poter essere tipizzato in campioni degradati. Negli ultimi anni sono state fatte interessanti identificazioni storiche grazie all'analisi del DNA mitocondriale.

Ad esempio, analizzando i resti umani ritrovati in una "Tomba del Milite Ignoto" della guerra del Vietnam, è stato possibile risalire all'identità del soldato ucciso in battaglia. Altro esempio, lo Zar Nicola II è stato identificato dalle ossa scoperte in Russia nel 1991. Tramite l'analisi del DNA mitocondriale si riuscì anche ad avere la certezza che i resti riesumati dalla presunta tomba del celebre criminale Jesse James erano effettivamente i suoi.

Il mitocondrio è un organulo cellulare di forma allungata, presente in tutti gli Eucarioti. È lungo 1-4 μm , ha un diametro di 0,2-1 μm ed è delimitato da una doppia membrana. Possiede al suo interno un piccolo cromosoma, in più copie, che si replica autonomamente. Nella cellula umana quasi tutti i geni (circa 30.000) sono confinati nel nucleo in duplice copia e sono trasmessi in parti uguali dai genitori secondo le leggi di Mendel.

I 37 geni del DNA mitocondriale (mtDNA) si trovano, invece, nei mitocondri, organizzati in una piccola molecola circolare chiusa di DNA a doppia elica costituita da 16.569 coppie di basi, che corrispondono allo 0,0006% del genoma totale, e codificano per due RNA ribosomiali (RNA), 22 RNA

di trasporto (tRNA) e 13 proteine che fanno parte dei complessi enzimatici deputati alla fosforilazione ossidativa.

La trasmissione del genoma mitocondriale avviene per linea materna. Infatti, i mitocondri molto abbondanti nel citoplasma degli ovociti sono gli unici conservati nello zigote dopo la fecondazione. Infatti i pochi mitocondri presenti nel collo dello spermatozoo non entrando nella cellula uovo non partecipano alla fecondazione.

Quindi, poiché il corredo genomico mitocondriale completo di tutte le cellule di ciascun individuo deriva dalla madre, attraverso il mtDNA si può indagare, solo la linea femminile delle genealogie, escludendo perciò il contributo maschile alla storia delle popolazioni in esame.

Dato che il processo di ricombinazione omologa non avviene nel DNA mitocondriale, le differenze che si accumulano per mutazioni casuali vengono trasmesse in blocco alle generazioni successive. Come già accennato la grande quantità di DNA mitocondriale presente nella cellula ne facilita l'analisi; i geni mitocondriali sono ripetuti migliaia di volte. Tale proprietà è particolarmente utile nell'analisi del DNA antico. Infatti con il passare del tempo il progressivo degradarsi dell'acido desossiribonucleico, permette uno studio significativo solo di quei geni inizialmente presenti in grandi quantità, come quelli mitocondriali appunto.

Probabilmente a causa della ridotta efficacia dei sistemi di riparazione, il DNA mitocondriale umano è caratterizzato da un elevato tasso evolutivo. Si stima che questa molecola subisca una o due mutazioni puntiformi ogni 100 nucleotidi per milione di anni; ciò gli permette di evolvere circa 10 volte più velocemente rispetto ai geni del DNA nucleare.

Di conseguenza, l'mtDNA viene assunto come un "orologio molecolare" che batte con un ticchettio molto veloce (in termini di mutazioni a suo carico), che si presta a documentare gli studi di antropologia molecolare e in generale l'evoluzione delle specie che si sono diversificate in epoca recente.

In particolare la regione di controllo (D-loop) non presenta geni codificanti ed è caratterizzata da due segmenti altamente variabili, a livello dei quali si osserva la maggior parte del polimorfismo presente nel DNA mitocondriale. Queste due aree sono identificate come la prima e la seconda regione ipervariabile, chiamate rispettivamente HVR-1 e HVR-2 (Figura 6).

codificata, dal momento che il codice genetico è degenerato. Uno SNP che genera in tutte le sue forme lo stesso peptide è detto sinonimo (synonymous); in caso contrario è detto non-sinonimo (non-synonymous). Gli SNP anche che non si trovano in una sequenza codificante possono produrre effetti negativi ad esempio sullo splicing o sul legame dei fattori di trascrizione.

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. Gli SNP hanno acquistato perciò, particolare rilevanza in campo biomedico quando sono stati 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 ed i fenotipi esibiti dai soggetti portatori, è possibile quindi utilizzare tali sequenze come marcatori molecolari.

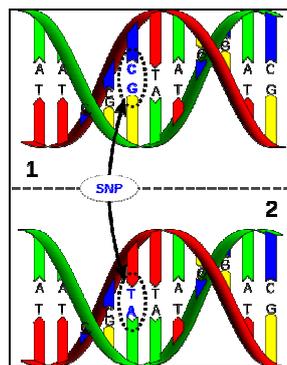


Figura 7: SNP

Identificazione di specie mediante l'analisi delle emoglobine

L'emoglobina è una proteina globulare solubile di struttura quaternaria, contenuta nei globuli rossi dei vertebrati. E' responsabile del trasporto dell'ossigeno molecolare da un compartimento ad alta concentrazione di O₂ come i polmoni ai tessuti periferici.

Per l'uomo sono note la struttura primaria, cioè la composizione della sequenza amminoacidica delle differenti catene globiniche; la secondaria, che riguarda la disposizione bidimensionale sul piano delle catene globiniche; la terziaria, riguardante le angolature e le ripiegature che le catene assumono nello spazio; e infine, la quaternaria relativa ai punti di contatto delle quattro catene globiniche fra di loro

disposte nello spazio. Le quattro catene polipeptidiche, a due a due identiche, costituiscono la porzione proteica (globina). Nel soggetto adulto sono presenti, in differenti proporzioni, tre tipi di emoglobina: emoglobina A, formata da due catene alfa e da due catene beta, che costituisce il 97% dell'emoglobina presente nei globuli rossi; l'emoglobina A2, costituita anch'essa da due catene alfa che però si combinano con due catene delta, e rappresenta il 2,5% del totale; infine, l'emoglobina fetale (HbF), 0,5%, costituita da due catene alfa e da due catene gamma (figura 8).

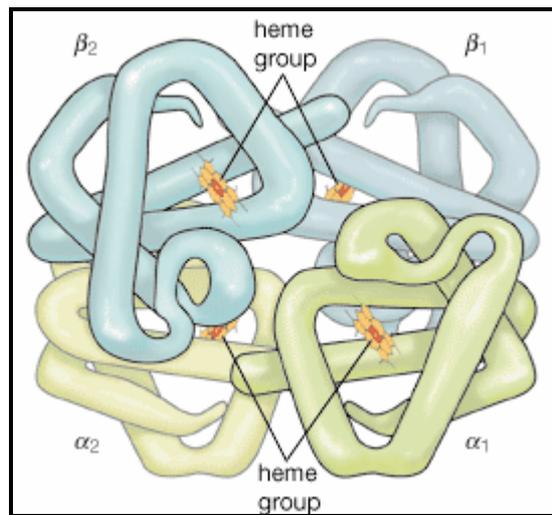


Figura 8: Emoglobina

Negli animali le catene globiniche differiscono dalle umane con una caratterizzazione non altrettanto dettagliata. La diversa nella lunghezza è una caratteristica propria delle differenti globine che le rende così ben discriminabile. Pertanto, utilizzando una metodica di separazione e rilevazione elettroforetica si possono identificare le diverse globine. La velocità e la precisione con cui si può effettuare questa analisi dipende dalla specifica metodica e dallo strumento utilizzato.

In campo forense, l'analisi per l'identificazione generica della presenza di sangue su un reperto viene effettuata mediante l'utilizzo del Combur Test, impiegato nell'area medica per l'analisi delle urine. Questo test è un saggio colorimetrico che serve a determinare il livello di glucosio, delle proteine e degli

eritrociti. Il tampone specifico per l'emoglobina di cui è composto reagisce colorandosi di verde in presenza dell'attività perossidasi di detta molecola (figura 9). Tale reazione essendo legata ad una proprietà chimico – fisica non esclusiva dell'emoglobina può dare falsi positivi, in particolare in presenza di perossidasi vegetali, ruggine o ossidanti di diverso tipo.

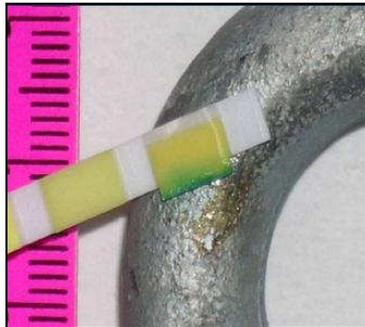


Figura 9: Combur test

Nel caso in cui il Combur Test risulta positivo, si procede ad una identificazione ancora più fine. Si cerca cioè di identificare se la traccia verosimilmente ematica possa esserlo effettivamente, ed in questo caso se sia umana o meno. Per fare ciò si mette in atto un test specie-specifico, come il FOBY o l'OBTI. Entrambi i test sono usati in ambito clinico per la diagnosi del sangue occulto nelle feci. Si tratta di saggi immuno-cromatografici basati sull'utilizzo di anticorpi anti-Hb (anti-emoglobina) umana coniugati a particelle cromogene blu che si legano con l'emoglobina umana garantendo una diagnosi di specie-specificità per il sangue di primati superiori (eventuali unici falsi positivi).

Quando il campione è inserito nella test bar, l'emoglobina presente reagisce con particelle blu e anticorpi monoclonali anti-hHb. L'immunocomplesso migra verso la zona test, dove è catturato da un secondo anticorpo immobilizzato diretto contro hHb formando una linea blu, indicando così un risultato positivo. Reagenti che non hanno reagito migrano ulteriormente e sono legati in una seconda linea da anticorpi IgG immobilizzati anti-mouse. Questa linea di controllo indica il corretto funzionamento ed utilizzo del test (figura 10).



Figura 10: OBTI test

Generalmente altre tecniche di identificazione sono applicate, tanto più per determinare la specie di appartenenza nel caso il materiale ematico non sia positivo ai test immuno-cromatografici. Eventuali tentativi di identificare la specie hanno coinvolto inizialmente l'utilizzo dell'HPLC sfruttando le differenti dimensioni delle globine nelle varie specie (Andrasko, 1994). Altri metodi, più moderni, si sono rivolti all'utilizzo di tecniche di genetica molecolare, analizzando la sequenza nucleotidica del citocromo b, che si trova nei mitocondri ed è specie-specifica (Parson, 2000). Queste tecniche molto sensibili e precise risultano però allo stesso tempo molto complicate, lunghe e costose, in confronto a tecniche di separazione elettroforetica.

SCOPO

La mia Tesi di Dottorato è stata sviluppata in laboratorio per approfondire lo studio dei polimorfismi allelici, per l'identificazione personale, e per investigare diverse emoglobine animali ed umane, per l'identificazione di specie. Per questi diversi campi di indagine ho impiegato la tecnica denominata "Denaturing High Performance Liquid Chromatography" o DHPLC. Questo strumento ha molteplici applicazioni nelle analisi in campo biologico. Durante lo svolgimento dei miei esperimenti ho applicato la DHPLC per caratterizzare delle specifiche varianti alleliche nelle regioni denominate D19S433, un esempio di marcatore specifico di DNA STR. Queste varianti dette microvarianti appaiono come alleli che si posizionano al di fuori dei ladder allelici convenzionali e che possono essere classificati come nuovi alleli mai individuati o come alleli rari. In una seconda fase per studiare le diverse globine appartenenti a distinte specie ho impiegato lo stesso strumento, DHPLC, a cui ho sostituito la colonna cromatografica per il DNA con una specifica per le proteine.

L'impiego della DHPLC è risultata di grande importanza per consentire un'indagine molto analitica e una fine tipizzazione genetica di micro satelliti, che come detto in precedenza, sono i marcatori molecolari più utilizzati in ambito forense. Il mio studio ha consentito lo sviluppo di una tecnica innovativa e rapida da affiancare al moderno sequenziamento del DNA che permetta di isolare e caratterizzare il solo allele selezionato.

Per la determinazione dei profili genotipici umani si utilizzano di norma due kit: l'"AmpFLSTR Identifiler Kit" della ditta Applied Biosystems, e il "PowerPlex 16" della ditta Promega. Questi due kit valutano il sesso del donatore analizzando il gene dell'Amelogenina e 15 loci, tra cui i 13 del CODIS ed altri 2 loci, che sono differenti per i due kit. Tecniche come il sequenziamento del DNA successivo ad una amplificazione multipla (multiplex), sono quindi già costantemente utilizzate nelle indagini forensi, ma l'analisi di campioni con uno strumento versatile come il DHPLC può risultare un'ottima prospettiva futura.

La possibilità di rilevare variazioni anche di un singolo nucleotide con il tipizzatore automatico ha permesso di identificare degli alleli che non sono costituiti da un multiplo intero delle unità ripetute che

costituiscono un STR. In questi casi, una di queste unità è presente solo in parte: tali alleli sono presenti in misura diversa all'interno delle popolazioni. Di rado può infatti capitare di trovare un amplificato di PCR mai osservato in precedenza, con una dimensione che corrisponde ad un numero di ripetizioni o parti di essa, non ancora pubblicato o noto. Questi amplificati vengono chiamati microvarianti. Pertanto, lo scopo della prima parte del mio lavoro è stato proprio quello di andare a verificare la formazione di una di queste microvarianti, la 6.2, sequenziando quattro di questi alleli identificati in una raccolta di quasi 30.000 campioni.

Dopo aver amplificato specificamente il marcatore D19S433 contenente il microsatellite raro, con la tecnica DHPLC, ho potuto separare i due alleli e raccogliarli in provette in condizioni adatte per poter essere utilizzati per altre analisi. Una parte della mia ricerca è stata svolta presso i laboratori del R.I.S. di Parma dove è presente un DHPLC in grado di separare nominalmente sequenze che differiscono per almeno due nucleotidi. Tale strumento, grazie ai due rilevatori (uno nello spettro UV-visibile ed uno in fluorescenza), consente di evidenziare anche minime quantità di DNA. Il suo utilizzo si è rivelato quindi di particolare utilità nell'identificazione di microvarianti alleliche, che rappresentano una ripetizione incompleta di un determinato allele e non sono presenti nel ladder allelico di riferimento.

Il riscontro di tracce organiche durante un sopralluogo può innescare una lunga serie di accertamenti che, in una sequenza a cascata, hanno come primo obiettivo l'identificazione della sostanza repertata e, successivamente, se possibile, l'analisi di quest'ultima per l'identificazione del soggetto da cui essa proviene. I limiti della ricerca e dell'accuratezza dell'indagine sono dati dalla quantità di materiale a disposizione e dallo stato di conservazione dello stesso. Ovviamente i limiti per l'identificazione della sostanza biologica differiscono rispetto a quelli per l'individuazione di un profilo genotipico; in questo caso infatti si lavora a concentrazioni inferiori ad 1 ng/μl (una cellula contiene circa 6 pg di DNA), mentre per l'identificazione della sostanza biologica le concentrazioni variano in base al test specifico utilizzato.

Nella seconda parte della mia ricerca ho analizzato emoglobine di specie diverse. Tale indagine è stata eseguita perché sulla scena del crimine di sovente sono presenti tracce di sangue non sempre ascrivibili

a sangue umano. Distinguere le diverse emoglobine rappresenta quindi, una indagine significativa per discriminare le diverse tracce ematiche tra di loro. La prima fase delle analisi prevede di identificare con certezza la natura della traccia in questione. Per quanto riguarda il sangue, esistono dei test diagnostici generici che sono basati principalmente sulle capacità ossido-reduttive dello ione Ferro presente nell'emoglobina, anche se il test in parola non è specie-specifico, poiché reagisce positivamente anche in presenza di sangue animale. Successivamente, qualora il test suddetto abbia dato esito positivo, si utilizza un test diagnostico specifico per l'emoglobina umana, di tipo immuno-cromatografico. Il risultato di quest'ultimo test non permette però di distinguere, nel caso si tratti di sangue animale non umano, la specie di appartenenza.

L'HPLC è una delle tecniche che permettono di separare le catene globiniche nell'emoglobina, ad esempio per distinguere l'Hb fetale dall'Hb adulta (Inoue et al., 1990) o per il riconoscimento di specie animali comuni e primati (Inoue et al., 1990).

Questo è il motivo per il quale ho messo a punto una metodica sensibile, affidabile, automatizzabile ed economica che permetta di identificare la presenza di materiale ematico e la specie di appartenenza, per confronto con i risultati dell'analisi di un pattern di animali tra i più diffusi nell'ambiente domestico. Per far ciò ho utilizzato la DHPLC sostituendo la colonna per il DNA con una appropriata per le analisi delle proteine. Con lo strumento così settato ho proceduto ad analizzare le globine presenti nell'emoglobine e a registrare i differenti tempi di ritenzione per le diverse specie animali. Questo metodo di analisi, inoltre, permette di non distruggere la traccia e di non disperdere il DNA che può così essere analizzato.

Lo scopo di questa parte di lavoro è stato appunto quello di identificare se questo metodo abbia le caratteristiche minime per essere considerato adatto all'utilizzo in alternativa ai sistemi ora diffusamente utilizzati. Per fare ciò dovevo valutare se avesse la sensibilità e la precisione minima indispensabile di identificazione del materiale ematico paragonabile ai test in utilizzo. Se fosse poi anche accurato ed in grado di identificare specificamente la specie di appartenenza, mediante confronto del campione in esame con un pattern di emoglobine animali analizzate a priori.

MATERIALI E METODI

Campioni

I campioni di DNA in analisi sono stati isolati da campioni biologici dei laboratori di biologia del R.I.S. di Parma. I DNA sono stati tipizzati per il locus D19S433, che mappa nella regione 12-13.1 del braccio lungo del cromosoma 19 del cariotipo umano (19q12-13.1).

I campioni di DNA provengono dall'archivio della struttura del R.I.S. di Parma dove sono presenti una notevole quantità di campioni biologici inclusi i DNA umani indagati da circa la metà degli anni novanta. Questa fonte di sequenze di DNA rappresenta un importante campione biologico a forte valenza scientifica.

I DNA da me scelti sono stati impiegati per l'analisi mediante la tecnica del DHPLC. Questo approccio mi ha consentito di analizzare questi genomi umani in condizioni non denaturanti, cioè con separazione basata sulla dimensione del frammento. I campioni sono stati codificati con un numero di serie per garantire l'anonimato con la conoscenza del genere.

I diversi campioni di DNA sono stati indagati per la sequenza del microsatellite D19S433. Di questi quattro campioni presentano la microvariante 6,2, quattro campioni hanno un allele normale e per ulteriore confronto ho analizzato un allele che presenta una ripetizione non completa. In particolare sono stati analizzati i seguenti campioni e i rispettivi allele indicati sotto:

≈ Campione 1: 6,2

≈ Campione 2: 6,2

≈ Campione 3: 14

≈ Campione 4: 12 – 14,2

≈ Campione 5: 6,2 – 14

≈ Campione 6: 6,2 – 13

PCR (Reazione a catena della polimerasi)

La reazione a catena della polimerasi, conosciuta come PCR nella sua sigla inglese, è una tecnica che consente di ottenere un gran numero di copie di un dato frammento di DNA, partendo da una quantità minima (in teoria, anche da un'unica molecola) di questo frammento. Questa tecnica si basa sulla proprietà naturale della DNA polimerasi termoresistente del batterio estremofilo *Thermus aquaticus* di duplicare il DNA. Questa caratteristica consente di impiegare uno strumento denominato termociclatore che in automatico, una volta settato, prevede per circa 35 cicli l'impiego di temperature che variano da 95° C., 55° C., 72° C. per: 1. separare i filamenti di DNA; 2. consentire l'appaiamento dei primers al DNA omologo denaturato; 3. sintesi del DNA (Figura 11).

Per la PCR si utilizzano oligonucleotidi sintetici di circa 20 nucleotidi, complementari alle zone fiancheggianti la regione che si vuole amplificare. Questi oligonucleotidi sono detti "primers" in quanto innescano la sintesi del DNA catalizzata dalla polimerasi. La Taq polimerasi è resistente alle alte temperature (95-97° C) ed è capace di sintetizzare DNA estendendo più di 60 nucleotidi per secondo in regioni ricche in GC.

La reazione a catena della polimerasi si basa sulla ripetizione di un ciclo formato da tre tappe (Mullis, 1990; Vosberg, 1989):

1. Denaturazione: la doppia elica del DNA si separa nei suoi due filamenti. Per far ciò, si esegue un'incubazione dei campioni ad alte temperature (90-95° C) che provoca la rottura dei ponti ad idrogeno: se la denaturazione si attua solo parzialmente, il DNA tenderà a rinaturare, impedendo l'ibridazione dei primers e la successiva estensione.
2. Ibridazione: è anche detta fase di annealing; si realizza grazie all'abbassamento della temperatura (40-60° C), quando i primers si uniscono alle zone 3' complementari a quelle che sono accanto al frammento che vogliamo amplificare. La T_m (melting temperature, o temperatura di fusione) dipende da vari fattori, soprattutto dalla grandezza e dalla proporzione di CG, ed è relativamente specifica per ogni primer. La formula per calcolare la T_m è:

$$T_m = 4(G + C) + 2(A + T)$$

Nonostante ciò, ogni primer ha bisogno di uno studio sperimentale per determinare la sua t° di fusione specifica, per far sì che l'unione di tutti i nucleotidi sia completa.

3. Estensione: si ha la sintesi di una catena semplice (si produce un frammento di doppia catena per la complementarietà) in direzione 5' - 3' grazie all'enzima DNA polimerasi, che incorpora i deossinucleotidi presenti nel campione seguendo la catena stampo.

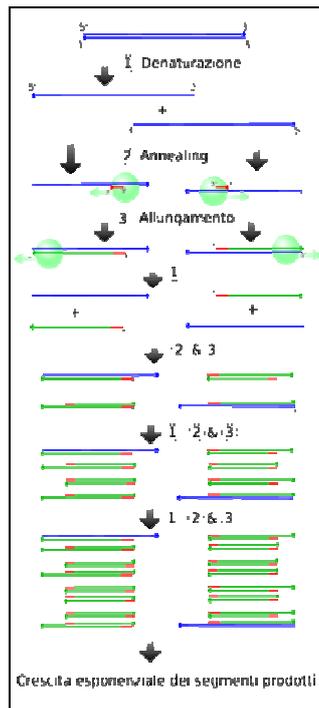


Figura 11: Reazione a catena della polimerasi.

Condizioni di amplificazione

Per l' amplificazione del locus D19S433 ho usato la seguente coppia di primers:

Forward 5'CCTGGGCAACAGAATAAGAT3'

Reverse 5'TAGGTTTTTAAGGAACAGGTGG3'

Le condizioni di amplificazione sono le seguenti:

- i. Attivazione della Taq-Gold: 95° C per 11 minuti
- ii. Ciclo di amplificazione (30 cicli):
 - a. Denaturazione..... 94° C per 1 minuto
 - b. Annealing..... 61° C per 1 minuto
 - c. Allungamento..... 72° C per 1 minuto
- iii. Fase finale di allungamento: 60° C per 1 ora e 20 minuti

Nel mio caso ho amplificato il DNA sia prima che dopo la separazione mediante DHPLC. La differenza tra la prima e la seconda amplificazione, effettuata per un singolo allele dopo corsa e raccolta delle frazioni con DHPLC, è la concentrazione dei primers; nella prima i primers erano ad una concentrazione di 100 µM, nella seconda la concentrazione era di 2 µM.

Multiplex

La PCR multiplex permette di amplificare simultaneamente più di una regione di DNA, aggiungendo semplicemente più set di primers specifici per le sequenze target. Affinché una reazione multipla di PCR fornisca risultati corretti, è necessario che le coppie di primers utilizzate siano compatibili; cioè le temperature di annealing degli oligonucleotidi devono essere simili e inoltre cercare di evitare un'eccessiva complementarietà tra primers per non permettere la formazione di dimeri tra i primers. In questa situazione il primer si unisce ad un altro primer invece di legarsi al DNA target.

Nel 1993 fu reso disponibile il primo kit di PCR Multiplex per gli STR. Questa applicazione si basa sull'amplificazione multipla, cioè in contemporanea, di più loci STR: tutto ciò per ridurre i tempi d'attesa ed aumentare il potere discriminante di una singola reazione di PCR. Il primo kit includeva 3 loci (TH01, TPOX CSF1PO) prodotto dalla Promega e con il passare degli anni i kit diventarono sempre più funzionali con l' aumento del numero di loci amplificabili. La prima PCR multipla applicata ad un caso forense riguardava l' amplificazione di 4 microsatelliti: il Match Probability ottenuto risultava troppo elevato (1/10.000), quindi per ottenere la sicurezza di individuare il colpevole fu

eseguita parallelamente un'analisi RFLP Single-locus Probe. Negli anni successivi si cercò di aggiungere a questi kits di amplificazione multipla un numero di loci microsatelliti tale da ottenere un Match Probability minore di 10^{-13} , per avere una quasi completa certezza che soltanto una persona nell'intera popolazione umana avesse quel determinato profilo genetico. La tecnica della PCR multipla risulta, quindi, essere molto informativa, riduce il lavoro manuale e riduce la quantità di DNA necessaria per ottenere un risultato utile e preciso. Nel 1996 l'FBI sponsorizzò un progetto, che coinvolse 22 laboratori di tipizzazione del DNA, per valutare quali loci STR includere in un grande database di sequenze di DNA conosciuto come CODIS (Combined DNA Index System) (Figura 12). Di 17 loci STR candidati ne furono scelti 13: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11. L'analisi contemporanea di questi loci determina un Match Probability talmente basso da escludere che due individui non imparentati possano avere lo stesso profilo genetico.

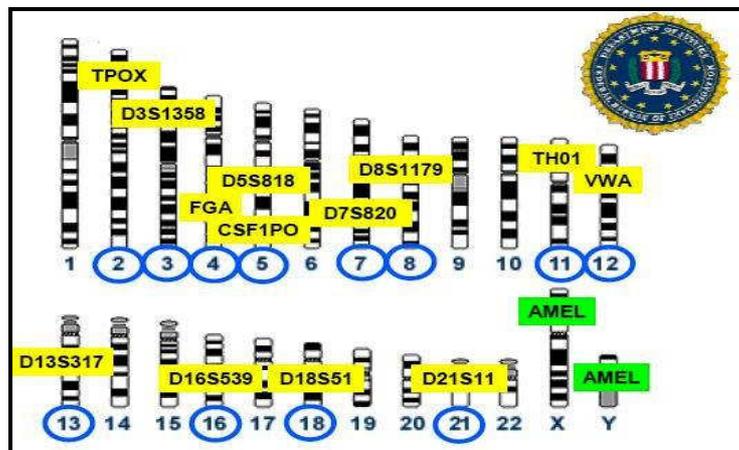


Figura 12: Elenco di marcatori microsatelliti utilizzati dal CODIS dell'F.B.I.

Negli ultimi anni sono stati messi in commercio due kit (Figura 13) che includono la co-amplificazione di 16 loci:

- PowerPlex 16 (Promega Corporation) con: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21D11, PENTA D, PENTA E ed Amelogenina.
- Identifiler (Applied Biosystem) con: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21D11, D2S1338, D19S433 e Amelogenina

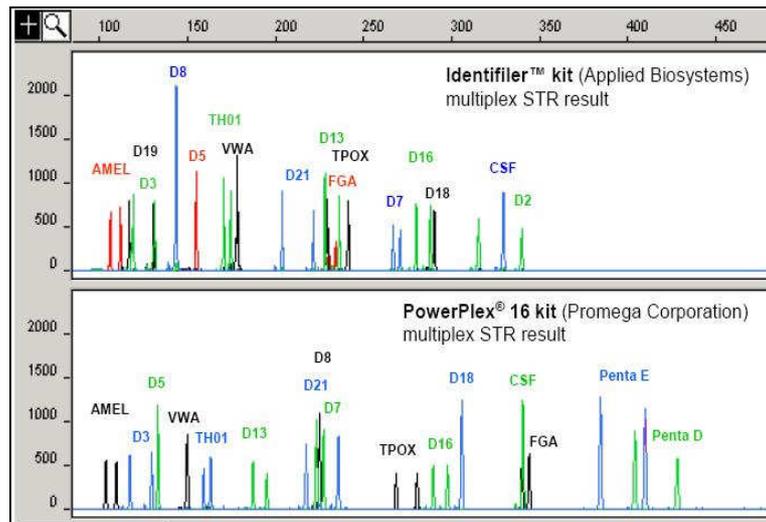


Figura 13: Esempio di due multiplex; la superiore è la rappresentazione grafica del kit Identifiler, quella inferiore del kit PowerPlex

Tutti i loci analizzati sono costituiti da un modulo di 4 basi nucleotidiche ad eccezione di PENTA D e PENTA E che sono di 5 basi nucleotidiche.

La capacità di predire se un campione di DNA appartiene ad un uomo o ad una donna risulta molto utile nei casi di abuso sessuale, infatti è importante in queste situazioni distinguere le tracce appartenenti alla vittima e quelle dell' aggressore. Allo stesso modo, investigazioni riguardanti persone scomparse o disastri di massa, possono trarre beneficio dall' identificazione sessuale ottenuta tramite l' analisi del DNA sui resti delle vittime.

Il sistema dell' Amelogenina permette la distinzione tra individui di sesso maschile e femminile.

L'Amelogenina è una proteina della matrice extracellulare espressa negli ameloblasti durante lo sviluppo dentale e coinvolta nel processo di produzione dello smalto. L'analisi del gene Amelogenina localizzato sulle braccia corte del cromosoma X e del cromosoma Y è in grado di distinguere la presenza dei due cromosomi sessuali (il gene dell'amelogenina è situato nella regione p22 del cromosoma X, mentre sul cromosoma Y è situato nella regione p11 e presenta un'inserzione di 6 bp). In caso di individuo di sesso femminile il locus Amelogenina è rappresentato sul tracciato elettroforetico da un solo picco (Figura 14), relativo al cromosoma X (XX), in caso di individuo di sesso maschile sono presenti due picchi (uno relativo al cromosoma X e l'altro all'Y).

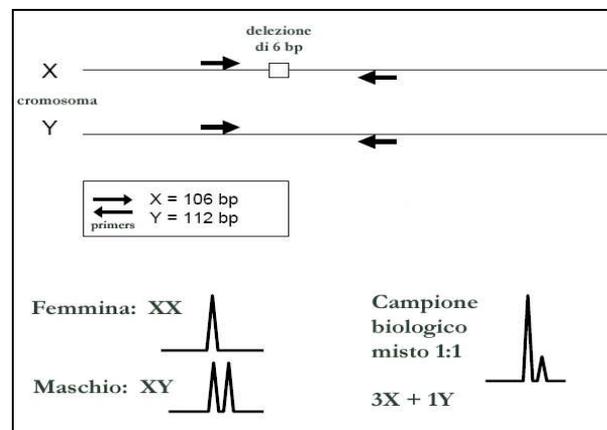


Figura 14: Rappresentazione del locus Amelogenina

MINELUTE Purification Kit (Qiagen)

Questo kit permette la purificazione diretta nelle reazioni di amplificazione di prodotti di PCR a doppia elica (70 pb – 4 Kb). Su un volume di eluizione di 10 µL, viene recuperato l'80% del DNA, ottenendo un eluato di circa 10 µL.

Il protocollo di purificazione del DNA permette di eliminare primers, nucleotidi, polimerasi e sali, usando le colonnine MinElute. Questo procedimento si suddivide in più fasi:

- Aggiungere 5 volumi di Buffer PBI (che favorisce il binding dei prodotti di PCR e la rimozione di primers) ad 1 volume di mix di reazione della PCR (es. 250 µL di Buffer per 50 µL di reaction mix).

Il colore di questa miscela deve essere giallo, ad indicare un corretto pH. Se il colore tende al rosso, il pH va corretto grazie all'aggiunta di Sodio Acetato.

- Spostare la miscela ottenuta in una colonnina MinElute e centrifugare a 17.900 g per 1 minuto.
- Dopo aver buttato l'eluato, rimettere la colonnina dello stesso tubo e aggiungere 750 μL di Buffer PE, ricentrifugando – sempre a 17.900 g – per un altro minuto. Una volta completata la centrifugazione, scartare l'eluato e ripetere la centrifugazione per 1 minuto. Questa seconda centrifugazione permette di eliminare tutte le tracce di Etanolo.
- Mettere la colonnina in una Eppendorf nuova e aggiungere 10 μL di Buffer EB (composto da 10 mM Tris-Cl a pH 8.5), lasciare riposare 1 minuto e centrifugare per 1 minuto a 17.900 g. in questo passaggio é importante che il buffer sia pipettato direttamente al centro della membrana, per permettere una completa eluizione del DNA.

BigDye Terminator Cycle Sequencing Kit (Applied Biosystems)

Nel BigDye Terminator Cycle Sequencing Kit i componenti necessari per la reazione di sequenziamento si trovano già mescolati in una miscela; l'unica cosa da aggiungere sono i primers ed il DNA. Questi reagenti sono adatti per reazioni di sequenziamento basate su fluorescenza del DNA a singola o doppia elica, ma anche su frammenti di reazioni di PCR, o ancora su DNA molto lungo (es. BAC, cloni...). Le concentrazioni dei reagenti per l'amplificazione effettuata con il BigDye Terminator Cycle Sequencing Kit, svolta sul Forward e sul Reverse di ogni allele separatamente, sono:

- ≈ 4 μL di BigDye
- ≈ 1,6 μL di primer
- ≈ 1 μL di DNA
- ≈ 13,4 μL di H_2O

La quantità assoluta di DNA deve essere di circa 10 ng, perciò un'eventuale diluizione del DNA da mettere in miscela verrà effettuata dopo aver quantificato ogni campione grazie ad una corsa in DHPLC.

Tale corsa avviene dopo il passaggio di purificazione con il MinElute, confrontando il picco di interesse con quello di un ladder standard (WAVE DNA Sizing Standard).

Le condizioni di amplificazione, su PCR Applied Biosystem 9700, sono state le seguenti:

- i. Attivazione della Taq-Hot Start:95° C per 11 minuti
- ii. Ciclo di amplificazione (32 cicli):
 - a. Denaturazione.....94° C per 1 minuto
 - b. Annealing..... 61° C per 1 minuto
 - c. Allungamento..... 72° C per 2 minuti
- iii. Fase finale di allungamento: 60° C per 10 minuti

DyeEX 2.0 spin (Qiagen)

Una volta ultimata la corsa in PCR con il BigDye, nella miscela vi è una grande quantità di dNTPs marcati: questi, durante il sequenziamento, possono causare “bolle” di colore a livello delle 30 pb e delle 70 pb, che devono essere eliminate per avere un profilo ottimale.

Il Kit DyeEx contiene colonnine che vengono usate per una veloce e facile rimozione dei dNTPs non incorporati; le colonnine contengono una resina pronta per essere usata, che permette di eliminare i dNTPs in soli 7 minuti.

Vi sono diversi step da eseguire: per prima cosa, la resina viene vortexata per essere risospesa, e successivamente viene centrifugata. La colonna contenente la resina viene poi trasferita molto attentamente in una eppendorf e viene aggiunto, esattamente nel centro della resina stessa senza toccarla, il DNA ottenuto dopo lo step con il BigDye (circa 10-20 µL). Dopo aver centrifugato, si ha il DNA pronto per il sequenziamento.

DHPLC (Denaturing High Performance Liquid Chromatography)

Il termine cromatografia deriva dal greco “krómatos” (colore) e “graphìa” (scrivere) che letteralmente significa scritto in colore. Le sue origini risalgono ai primi anni del XX secolo (1901-1903), quando il

botanico russo Michail Semenovitch Tswett filtrò una soluzione di sostanze vegetali su di una colonna riempita con del calcio carbonato (come materiale adsorbente) ottenendo una separazione dei pigmenti colorati. Da tempo, la cromatografia, nelle sue varie realizzazioni, si è dimostrata indubbiamente una delle più efficaci e sfruttate tecniche per la risoluzione dei difficili problemi che si presentano in campo analitico ai ricercatori. Tutti i tipi di cromatografia sono basati sul fatto che, ogni composto di una miscela interagisce con l' ambiente circostante differentemente dagli altri, nelle stesse condizioni. Ogni sistema di separazione viene determinato dalla distribuzione competitiva di un componente tra una fase mobile (fluido vettore o eluente) e una fase stazionaria; attraverso il supporto viene realizzata, in base a determinati parametri, una migrazione differenziale delle diverse sostanze e quindi la loro separazione. Logicamente, non si ha separazione se i componenti non sono almeno parzialmente trattenuti dal materiale della colonna; questo adsorbimento è determinato dal numero e dal tipo di gruppi chimici funzionali presenti.

Utilizzando al meglio le caratteristiche delle fasi e le specificità delle singole molecole è generalmente possibile risolvere una miscela di componenti sia in modo qualitativo che quantitativo.

La DHPLC può essere impiegata come tecnica per isolare frammenti di DNA, per la rilevazione di nuove mutazioni (SNP, inserzioni, delezioni e tandem repeat) o per lo screening su popolazioni per la ricerca di una specifica mutazione e per l'identificazione di malattie (Xiao e Oefner, 2001). Può essere utilizzata anche per purificare i prodotti provenienti dalla PCR, per garantirne un uso esente da contaminazioni, e per separazioni di frammenti di DNA tagliati con enzimi di restrizione. A parte questi usi, la DHPLC è stata utilizzata per la mappatura e la clonazione di geni di lievito (Steinmetz et al., 2002), *Arabidopsis thaliana* (Spiegelman et al., 2000), moscerino della frutta (Hoskins et al., 2001) e topo (Schriml e Dean, 2000); inoltre, è stata utilizzata per la misura dell'espressione genica (Doris et al., 1998) e dell'analisi di prodotti dell'estensione di singoli nucleotidi (Hoogendoorn et al., 1999).

La DHPLC è una tecnica sviluppata nel laboratorio del Prof. Cavalli-Sforza alla Stanford University per la rivelazione di mutazioni nel DNA.

Questa tecnica basa la sua capacità discriminante per la diversa velocità di migrazione degli eteroduplex e degli omoduplex. Questi duplex si formano quando un frammento amplificato di DNA mutato ed uno

non mutato vengono denaturati termicamente e lasciati appaiare a temperatura ambiente. Una qualsiasi variazione tra la molecola originale (wild type) e quella mutata porta alla formazione di un eteroduplex (appaiamento di due singoli filamenti di DNA non perfettamente corrispondenti, caratterizzata perciò dalla presenza di una "bolla" dovuta di non appaiamento nel punto dove è avvenuta la mutazione – Figura 15).

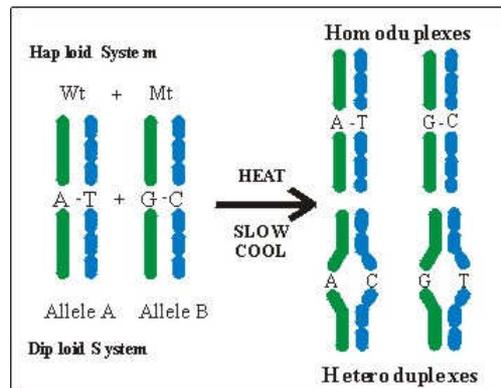


Figura 15: omoduplex ed eteroduplex.

L'eteroduplex si comporta cromatograficamente in modo diverso dall'omoduplex: i tempi di ritenzione sono infatti differenti, l'eteroduplex è solitamente più veloce dell'omoduplex e da ciò si può caratterizzare la presenza di una mutazione in un campione.

La DHPLC rileva la differenza tra la molecola dell'omoduplex - che ad una determinata temperatura è ancora sotto forma di due filamenti - e quella dell'eteroduplex - che alla stessa temperatura mostra una parziale denaturazione in corrispondenza del sito dove si è verificato un mis-appaiamento (Premstaller e Oefner, 2002).

La presenza di una mutazione si evidenzia sotto forma di picchi ulteriori rispetto al "wild type"; il grande vantaggio è che - pur non caratterizzando la mutazione (cioè non viene definito come è stata modificata la sequenza) - la DHPLC è in grado di rivelarne la presenza all'interno del frammento analizzato.

La DHPLC impiega il meccanismo di ripartizione in fase inversa ad accoppiamento ionico (RP-IPHPLC), che separa le molecole e le eluisce sulla base della dimensione crescente delle molecole.

Una molecola viene trattenuta dalla fase stazionaria o trasportata dalla fase mobile in dipendenza del suo coefficiente di ripartizione (CR). La ripartizione di un componente nelle due fasi viene descritta dal suo coefficiente di ripartizione o di distribuzione che è dato dal rapporto tra la concentrazione del componente in una delle due fasi e la concentrazione del componente nell'altra fase:

$$K_d = [\text{componente}]_{\text{nella fase1}} / [\text{componente}]_{\text{nella fase2}}$$

Il valore di tale coefficiente ad una determinata temperatura è costante e dipende dalla natura chimica del sistema, inteso come l'insieme delle due fasi e del componente stesso.

Variando nel tempo la concentrazione della componente organica nella fase mobile possono essere eluite in tempi diversi molecole con differenti CR.

I diversi elementi che compongono l'apparecchiatura per la DHPLC comprendono (vedi Figura 16):

- piastra sulla quale possono essere posizionati fino a 96 campioni;
- autocampionatore, un sistema di iniezione che permette di avere una grande precisione nel volume analizzato; inoltre, consente di rendere estremamente automatizzata una procedura che richiederebbe la costante presenza di un operatore addetto all'iniezione del campione;
- pompa che assicura un flusso costante attraverso la colonna di separazione cromatografica. Serve a fornire la pressione adatta per far scorrere gli eluenti attraverso la fase stazionaria; la stabilità del flusso è una delle caratteristiche più importanti che deve possedere una buona pompa. Grazie a queste pompe ad alta pressione, che garantiscono una portata costante e riproducibile, la dHPLC riesce a controllare e mantenere costante il flusso della fase mobile. Quanto minore è il diametro delle particelle che compongono la colonna, tanto maggiore è la pressione richiesta;
- colonna cromatografica, a forma tubolare. Parametri fondamentali sono il suo diametro interno e la sua lunghezza. La maggior parte delle colonne è costituita da un supporto solido di particelle di silice;

- sensore spettrofotometrico di rivelamento che trasforma il segnale analogico in un valore digitale; i detector fanno passare un fascio di luce attraverso l'eluente, mentre questo scorre lungo la colonna e le variazioni nell'intensità della luce dovute all'assorbimento UV o all'emissione fluorescente vengono monitorate. Si ottiene così un cromatogramma in cui per ciascun soluto presente viene indicato il tempo di ritenzione e l'area del picco. Esistono più tipi di detector, la cui scelta si determina in base alle caratteristiche della sostanza che deve essere analizzata.
- sistema HSX, che permette di aumentare la sensibilità del segnale;
- rilevatore di fluorescenza, che è uno strumento opzionale; aumenta in maniera evidente la sensibilità e la versatilità del DHPLC. E' collegato al software per generare sul display in tempo reale il cromatogramma del campione analizzato;
- raccogliatore di frazioni, che, come lo strumento precedente, è uno opzionale. Lavora assieme al Software Navigator e al Sistema WAVE per raccogliere specifici picchi di DNA evidenziati in un cromatogramma. Risulta utile per analizzare e purificare oligonucleotidi, ampliconi di PCR o frammenti di acido nucleico non identificati, che possono essere eventualmente sequenziati o clonati dopo una successiva re-amplificazione.



Figura 16: La strumentazione del DHPLC

DHPLC WAVE® TRANSGENOMIC

Il "Sistema WAVE®" è uno strumento di analisi di frammenti di acidi nucleici umani, animali, vegetali e batterici completamente automatizzato, utile per le analisi con finalità di ricerca e per diagnosi di laboratorio. E' in grado di offrire diverse tipologie di configurazioni per ottenere una separazione automatica di frammenti di acidi nucleici a singolo o doppio filamento.

La colonna cromatografica (Figura 17) contiene una fase stazionaria (o matrice inerte), formata da granuli di gel con pori di dimensione controllata, in questo caso da particelle di polistirene-divinilbenzene (PS-DVB) con un diametro medio di 2,3 μm ; si tratta di una specie di setaccio molecolare in cui il diametro dei pori dipende dal numero di legami trasversali presenti nel materiale della matrice. Per natura, le sfere sono elettrostaticamente neutre ed idrofobiche, e non reagiscono con gli acidi nucleici.



Figura 17: Colonna cromatografica alla base del funzionamento del DHPLC

La fase mobile fluisce attraverso la fase stazionaria ed è costituita da un liquido, le cui caratteristiche vengono variate nel tempo. Questo liquido trascina con sé i componenti della miscela che più risultano affini ad esso. Il trietilammonioacetato agisce come "molecola ponte" per permettere l'adesione degli acidi nucleici alle sfere; questa molecola ha una porzione idrofobica ed una porzione carica positivamente. La carica positiva del TEAA interagisce con la carica negativa dei gruppi fosfato degli acidi nucleici, mentre i gruppi idrofobici interagiscono con le porzioni idrofobiche delle sfere di PS-

DVB presenti nella colonna (Figura 18). Le sfere sono alchilate con catene C-18 che formano singoli legami C-C.

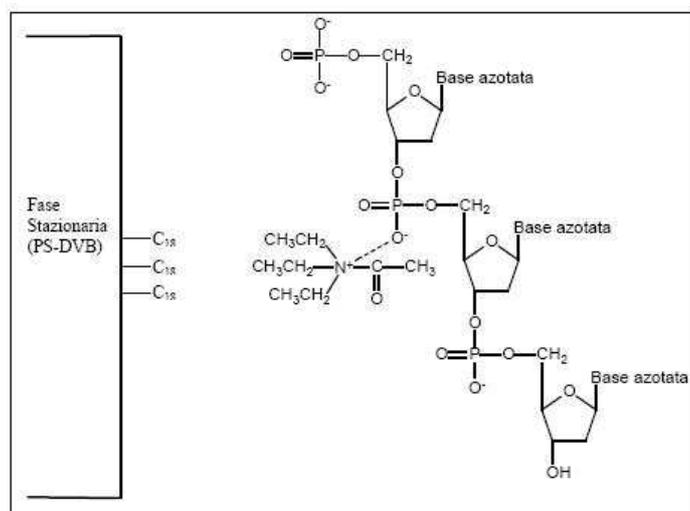


Figura 18: interazione fra FASE STAZIONARIA - TEAA – DNA

Una sostanza più affine alla fase stazionaria che alla fase mobile impiega un tempo maggiore a percorrere la colonna cromatografica rispetto ad una sostanza con bassa affinità per la fase stazionaria ed alta per la fase mobile.

In questo caso la fase mobile è una miscela di due soluzioni A e B. La soluzione A consiste in un reagente carico positivamente, il trietilammonioacetato (TEAA) 100 mM pH 7,0, mentre la soluzione B contiene TEAA 100mM e 25% di acetonitrile (ACN) pH 7,0. Il TEAA, costituito da una porzione idrofobica ed una idrofilica, permette alle molecole di DNA, cariche negativamente, di interagire con la matrice idrofobica della colonna facendo da ponte; la porzione idrofilica (carica positivamente) interagisce con i gruppi fosfato (carichi negativamente) delle molecole di DNA, mentre quella idrofobica interagisce con la fase stazionaria della colonna.

Questo strumento risulta molto versatile e può operare in tre modi differenti correlati ad una variazione di temperatura (Tabella 1):

- ✓ Condizione non denaturante
- ✓ Parzialmente denaturante

✓ Completamente denaturante

In tutti i casi il risultato è registrato in un cromatogramma, cioè un grafico prodotto da un'analisi cromatografica che correla la risposta del rivelatore del DHPLC al tempo.

Il successo di una separazione cromatografica è giudicato in base alla capacità del sistema di separare il picco di ogni analita presente nel campione da ogni altro.

Tipo di Metodo	Range di temperatura	Possibili Applicazioni	Base di separazione
Condizioni non denaturanti	~ 50° C	<ul style="list-style-type: none">• Sizing di DNA a doppio filamento (fino a 2000 bp)• Purificazione e controllo della qualità della PCR• Analisi quantitative (Q-RT-PCR)	Dipendente dalla dimensione Indipendente dalla sequenza
Condizioni parzialmente denaturanti	52-72° C	<ul style="list-style-type: none">• Ricerca mutazioni• Scoperta di SNP	Dipendente dalla dimensione Dipendente dalla sequenza
Condizioni completamente denaturanti	75-80° C	<ul style="list-style-type: none">• Sizing di DNA a doppio filamento (fino a 2000 bp)• Analisi di RNA• Analisi di oligonucleoidi	Dipendente dalla dimensione Dipendente dalla sequenza

Tabella 1: Possibili applicazioni del DHPLC a temperature variabili

Condizioni non denaturanti: separazione basata sulla dimensione

Il Software Navigator è in grado di creare metodi di separazione ottimizzati per separare frammenti di acidi nucleici in relazione alla loro dimensione. In queste condizioni, le caratteristiche della sequenza non sono determinanti nel comportamento che assume il DNA nella eluizione attraverso la colonna cromatografica. Il software programma l'analisi del campione ad una temperatura di circa 50° C e soltanto il numero di paia di basi del frammento determina il tempo di ritenzione nella colonna. Utilizzando questa applicazione si possono separare ed individuare inserzioni e delezioni nella sequenza di acido nucleico analizzata. Generalmente questo strumento permette di distinguere frammenti che differiscono per l' 1% della lunghezza. Per esempio, un prodotto di 100 bp può essere distinto da un frammento di 101 bp, così come una sequenza di 300 bp sarà ben discriminata rispetto ad una di 303 bp. I frammenti di DNA (o RNA) più corti hanno un numero minore di gruppi fosfato (carichi negativamente) che si legano alla matrice della colonna (con una complessiva carica positiva), mentre i frammenti più lunghi con più gruppi fosfato creano più legami. Questo metodo prevede un graduale aumento nel tempo della concentrazione di Acetonitrile (aumenta il flusso del tampone B) che fa diminuire la capacità del TEAA di creare legami ionici con l' acido nucleico; l'ACN ha infatti più affinità per il TEAA rispetto all'acido desossiribonucleico e fa rilasciare i frammenti di DNA. Per questo motivo i frammenti di acido più corti eluiscono dalla matrice della colonna più rapidamente e quindi prima di quelli di dimensioni maggiori (Figura 19). In questo caso si può, quindi, ottenere un'accurata analisi di lunghezza di frammenti di DNA.

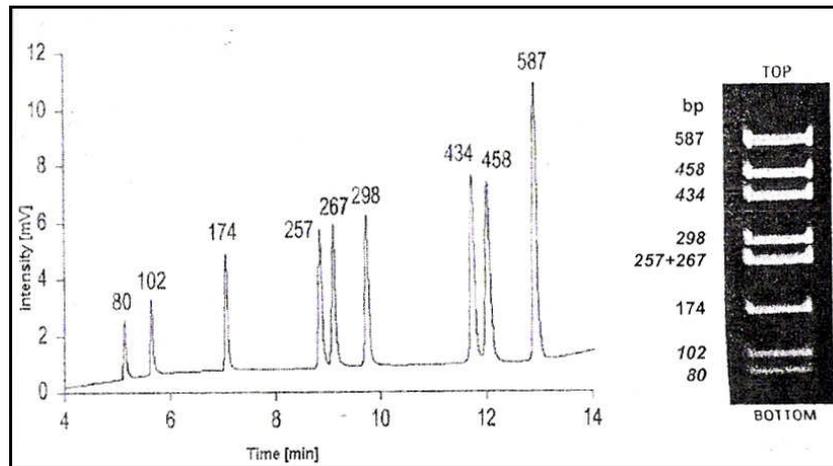


Figura 19: Cromatogramma di sizing di frammenti di DNA

Condizioni parzialmente denaturanti: separazione basata su dimensione, sequenza e temperatura

Il Software Navigator è anche in grado di predire le variabili ottimali per una analisi di ricerca di mutazioni in un frammento di DNA. In condizioni parzialmente denaturanti, i risultati di separazione chimica che si ottengono sono legati alla lunghezza, alle caratteristiche della sequenza e alla temperatura di analisi. Dopo un'amplificazione del DNA con una PCR, si può osservare la presenza o meno di una mutazione in una determinata regione del frammento.

Il protocollo prevede l'amplificazione separata dell'allele "normale" e del presunto allele mutato; successivamente si procede con il mix dei due amplificati in una unica provetta. Il metodo, a questo punto, prevede di portare il campione ad una temperatura di circa 95°C, con un successivo lento raffreddamento. La denaturazione dovuta all'elevata temperatura porta i frammenti di DNA allo stato di singolo filamento; con il raffreddamento, si osserva un riappaiamento delle sequenze: da questa situazione si possono generare, se la mutazione è presente, 4 possibili configurazioni: omoduplex wild type, omoduplex mutanti e 2 varianti di eteroduplex. Il DNA eteroduplex si genera quando un frammento amplificato di DNA mutato ed uno non mutato vengono a contatto, l'omoduplex è invece costituito dall'appaiamento di due frammenti identici. Si formano in tal modo le coppie di omoduplex e di eteroduplex che vengono iniettate dentro una colonna di cromatografia denaturante.

Il Software Navigator crea un metodo che controlla la temperatura del campione e il gradiente di eluizione di Acetonitrile. Mentre la molecola di omoduplex è intatta, la coppia eteroduplex mostra una parziale denaturazione nel sito di non perfetto appaiamento dovuto alla presenza della mutazione. Gli eteroduplex, con misappaiamento di coppie di basi, eluiscono dalla colonna prima rispetto agli omoduplex, perché l'eteroduplex ha una capacità di legare il TEAA inferiore ad una molecola omoduplex (Figura 20). Ciò è dovuto al fatto che nel punto della mutazione si formano delle porzioni di DNA a singola elica (dette bolle), poiché le due singole eliche non sono in grado di formare legami a idrogeno in quanto le basi risultano non correttamente appaiate.

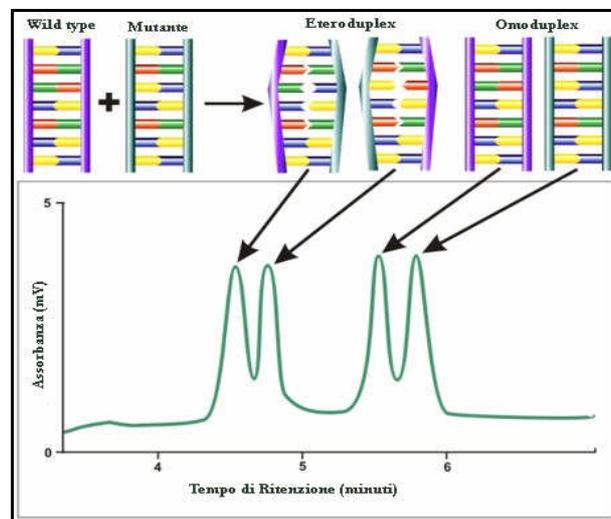


Figura 20: Cromatogramma che evidenzia la presenza di una mutazione in un frammento di DNA

La proprietà fondamentale di tali bolle, nel DNA a doppia elica (dsDNA), è quella di avere una minore densità di gruppi fosfato e, quindi, una minore densità di carica rispetto al doppio filamento, garantendo una minore ritenzione della molecola di acido deossiribonucleico sulla fase stazionaria e un tempo di eluizione inferiore a quello dell' omoduplex (Figura 21).

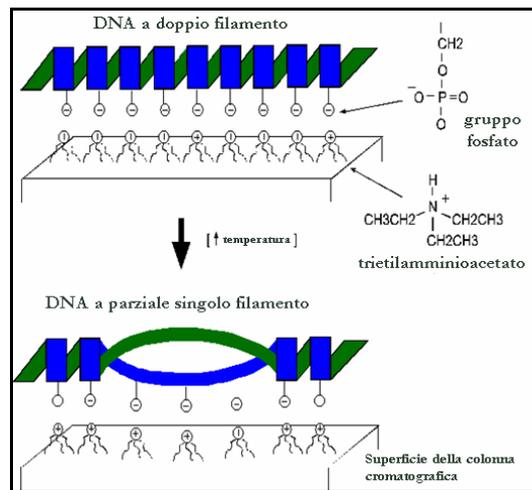


Figura 21: Variabilità del tempo di ritenzione in frammenti di DNA

Questo metodo risulta utile per effettuare screening di polimorfismi di singolo nucleotide (SNPs), quindi fondamentale nella ricerca di mutazioni puntiformi eventualmente correlate a particolari patologie; è questa la modalità di utilizzo più comune del DHPLC, con un importante valore diagnostico.

Condizioni completamente denaturanti: analisi di frammenti a singola elica basata su dimensione e sequenza

Il Sistema WAVE è in grado di analizzare oligonucleotidi (primers) e RNA utilizzando condizioni completamente denaturanti, cioè a temperature molto elevate. In questa situazione gli acidi nucleici vengono denaturati e separati sia sulla base della lunghezza della sequenza che sulle caratteristiche chimico-fisiche del frammento.

Dall'immagine sottostante (Figura 22) si può capire come un segmento di 16 paia di basi nucleotidiche fluisce successivamente rispetto a segmenti nucleotidici più brevi e viene chiaramente discriminato rispetto a quelli che differiscono di sola una base.

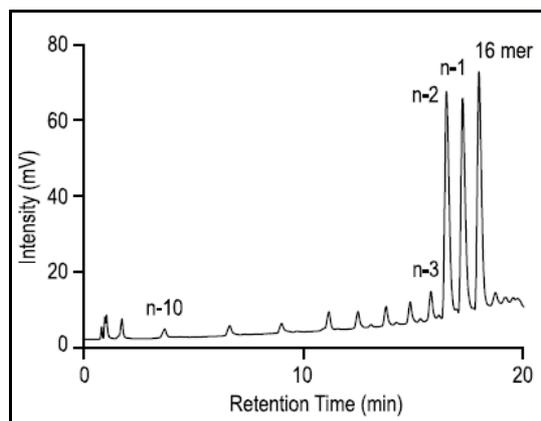


Figura 22: Cromatogramma di analisi di oligonucleotidi

Questo metodo permette di analizzare e purificare acidi nucleici a singolo filamento.

La strumentazione del DHPLC in queste analisi è essenzialmente uguale a quella usata per l'analisi delle microvarianti nel DNA. Le uniche differenze si riscontrano nelle lunghezze d'onda di eccitazione ed emissione della fluorescenza – rispettivamente 275 nm e 345 nm, e nella lunghezza d'onda dell'UV, che è 220 nm. Un'altra differenza si ha nel tipo di fasi utilizzate in queste analisi:

≈ Fase A: 0,1% TFA in acqua pentadistillata + 0,1% TFA in Acetonitrile in proporzione 80:20

≈ Fase B: 0,1% TFA in acqua pentadistillata + 0,1% TFA in Acetonitrile in proporzione 40:60

che vengono miscelate partendo da 40% di fase B fino ad arrivare al 60% di fase B, in 20 minuti.

La colonna è, in questo caso, una colonna C5 da HPLC 25 cm x 4.6 mm, 5 µm di particle size, della ditta Supelco.

Per testare la ripetibilità del metodo, le analisi sono state svolte utilizzando lo stesso metodo su un identico materiale (in questo caso sangue umano diluito 1:300) più volte nella stessa giornata (5 volte consecutivamente) e più volte in giorni differenti (3 volte al giorno per 3 giorni diversi) per verificare la precisione nelle analisi intra e inter-giornaliera.

Nello studio del metodo per la separazione delle globine si è individuata anche la sensibilità del metodo stesso. Per far ciò, le diverse analisi sono state effettuate con sangue umano fresco a concentrazioni differenti a partire da una concentrazione di 1:10 fino al limite minimo, che è risultato essere la concentrazione di 1:40.000.

Elettroforesi capillare

L'elettroforesi capillare si basa sugli stessi principi delle tecniche elettroforetiche convenzionali (separa frammenti per grandezza e per carica) ma usa condizioni e tecnologia diversa, che ci permettono di avere vantaggi rispetto alla tradizionale. Il supporto (o mezzo di separazione) è un polimero incluso in un capillare di silicone di 50 µm e di lunghezza variabile. Ad un lato e all'altro del capillare vi sono gli elettrodi che vanno a creare la differenza di potenziale che genererà il movimento dei frammenti del DNA. Il capillare dove si mette in atto l'elettroforesi è ricoperto da poliamide opaca, eccetto in una zona detta "finestra del capillare", che è la zona di attraversamento del laser. Il DNA deve essere amplificato usando un paio di primers, o più, marcati all'estremo 5' con delle molecole che emettano fluorescenza a diverse lunghezze d'onda quando eccitate dal laser (fluorocromi). Il software di recupero dei dati definisce particolari aree in una camera CDD per recuperare esclusivamente le emissioni provenienti dal gruppo di fluorocromi che stiamo usando (queste aree si chiamano filtri virtuali). Alla fine, i dati vengono inviati al pc che li trasformerà in sequenze di DNA o in frammenti con i loro corrispondenti alleli assegnati.

Il vantaggio della rapidità di questa tecnica è dovuto al fatto che la preparazione dei campioni e la loro caricatura vengono attuate automaticamente; in più, i risultati ottenuti sono analizzati da un software, evitando così problemi di interpretazione.

ABI Prism® 3130

I campioni da analizzare sono stati sequenziati in un sequenziatore genetico a 16 capillari ABI Prism® 3130 della Applied Biosystem (Figura 23), un sistema di elettroforesi capillare automatizzato basato sulla fluorescenza. Questo sistema di analisi fornisce dei risultati ottimali con una elevata capacità di

analisi, una corsa di 16 campioni dura circa 50 minuti. La tecnica dell'elettroforesi capillare offre molti vantaggi per l'analisi del DNA rispetto ai classici gel. Innanzitutto la temperatura è controllata permettendo una corsa maggiormente controllata, e l'iniezione elettrocinetica permette la simultanea iniezione di 16 campioni in circa 30 secondi mantenendo le dimensioni del campione molto limitate cioè fino a 10 microlitri. L'analizzatore del laser permette di effettuare analisi in 4 o 5 colori utilizzando matrici di selezione del picco di emissione. Per il funzionamento è indispensabile utilizzare il Sequencing Analysis Software che permette la gestione dello strumento. Per l'analisi degli STR ho utilizzato poi un software denominato GeneMapper Software con il quale si possono fare diverse analisi: genotipizzazione, chiamate alleliche, dimensionamento di frammenti di DNA e l'analisi degli SNP. Questo software nell'effettuare l'analisi assegna un valore di qualità alla singola corsa elettroforetica. Nel caso degli STR la corsa elettroforetica di un campione viene automaticamente confrontata con un ladder allelico, ed il software come risposta assegna ad ogni picco una chiamata allelica. L'operatore può a questo punto interagire e correggere il profilo così ottenuto.

Per il sequenziamento ho utilizzato il SeqScape Software che ha la possibilità di effettuare anche analisi di mutazioni. Questo software permette di inserire una sequenza di riferimento, di analizzare le corse elettroforetiche del DNA dopo la reazione di sequenziamento e di confrontarle tra loro. In questo modo effettua l'allineamento e individua eventuali mutazioni segnalandole. In questo caso l'operatore ha la possibilità di modificare la lettura del software dall'osservazione diretta della corsa elettroforetica.



Figura 23: ABI Prism 3130

RISULTATI

La microvariante 6.2 del locus D19S433 è presente nella popolazione mondiale con una frequenza calcolata di circa 1/11.500 persone prese a caso, come indicato dal sito del NIST (National Institute of Standard and Technology). In questo lavoro ho analizzato nove diversi alleli, identificati con il kit Identifiler ed il sequenziatore automatico di DNA. Ho potuto verificare che quattro alleli appartenevano alla forma 6.2 (e quindi della forma microvariante), mentre cinque alleli con vari polimorfismi, incluso uno denominato 14.2 che presentava una ripetizione incompleta. Per confronto ho considerato la sequenza presente in banca dati del NCBI, l'allele 13. Ho potuto verificare che la tipologia dei campioni di DNA disponibili, in quanto di tipo forense, era perlopiù di ridotte quantità e in parte degradato. Queste caratteristiche del campione hanno reso molto difficile la messa a punto del sistema analitico da me impiegato. Tuttavia superati i problemi tecnici ho potuto identificare l'allele di interesse. In seguito ho amplificato di quel dato campione i due alleli. Subito dopo, l'amplificato è stato analizzato cercando di ottenere le condizioni che permettessero la maggiore separazione possibile dei due alleli con il DHPLC. Solo allora ho effettuato la separazione con la raccolta delle frazioni a intervalli successivi, in modo da ottenere nel corso dell'eluizione un incremento progressivo della prima forma allelica a cui succedeva un calo ed un incremento nella seconda forma allelica. Le frazioni che venivano raccolte approssimativamente in corrispondenza del picco dell'allele prescelto sono state quindi riamplicate, e separate nuovamente con il DHPLC in modo da selezionare quella che presentava la quantità di DNA maggiore dell'allele considerato, e allo stesso tempo la minore quantità possibile del secondo allele. Detta situazione spesso non corrispondeva a quella in cui l'allele voluto era nella concentrazione maggiore. Una volta scelta la frazione migliore questa veniva pulita dai primer rimasti nella miscela di amplificazione e non utilizzati, attraverso il kit Minelute Purification Kit della ditta Qiagen. Successivamente gli amplificati purificati sono stati quantificati in maniera indiretta, confrontando l'area sottesa al picco dell'allele con l'area sottesa ad un picco di simili dimensioni in copie di basi nucleotidiche e facente parte del marker allelico fornito dalla ditta Transgenomic, di cui si conosce la concentrazione di DNA. A questo punto con il Sequencing Kit ho effettuato le reazioni di sequenziamento. Tutti i Dye in eccesso, che rendono non leggibile la sequenza devono essere tolti e per

fare ciò è stato utilizzato il DyeEX 2.0 Spin. Solo a questo punto ho caricato i processati nel sequenziatore automatico per la lettura delle sequenze utilizzando il software SeqScape come in esempio nella figura 24, nella quale sono mostrate le ripetizioni di un allele 6.2 con la delezione che porta al numero incompleto di ripetizioni.

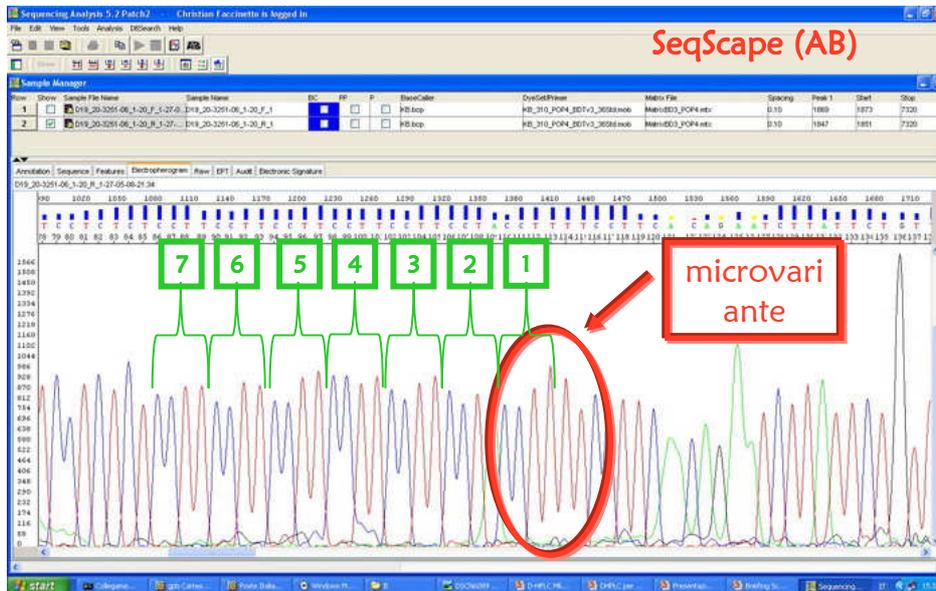


Figura 24: Sequenza della microvariante nel punto della delezione

L'analisi del DNA è stata eseguita per gli alleli considerati e tabulata confrontandola con la sequenza dell'allele 13 (NCBI). Nella figura 25 è possibile notare come la delezione di due basi AG nella posizione 63 e 64 della sequenza G08036 composta di 269 pb del sito NCBI sia quella che sta alla base di tutte le forme di ripetizioni parziali nonché delle microvarianti. La posizione è la medesima di quella riscontrata negli alleli maggiormente diffusi e cioè 10 pb prima dell'inizio della sequenza ripetuta $(AAGG)_n$. Oltre alla mutazione già specificata si possono individuare delle altre Transizioni non collegabili alla rarità dell'allele che le presenta.

6,2	1	T	T	T	C	T	C	T	G	G	G	C	A	C	A	G	A	A	T	A	A	G	A	T	T	
6,2	2	T	T	T	T	C	T	G	G	G	C	A	C	A	G	A	A	T	A	A	G	A	T	T	T	
14	3	T	T	T	T	C	T	G	G	G	C	A	C	A	G	A	A	T	A	A	G	A	T	T	T	
12	4	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
14,2	5	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
6,2	6	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
14	7	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
6,2	8	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
13	9	G	C	A	C	T	C	C	A	G	C	A	C	T	C	C	A	G	C	A	C	T	C	C	A	G
D19 Sequenza NCBI (13)		C A A A G C T A T A A T T G T A C C A C T G C A C T C C A G																								

- Primer Forward
- Inserzione
- Transizione (A/T - C/G)
- Primer Reverse
- Delezione
- Transversioni (C/T - A/G)

6,2	1	A	G	A	A	G	A	A	G	A	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
6,2	2	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
14	3	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
12	4	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
14,2	5	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
6,2	6	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
14	7	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
6,2	8	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
13	9	A	G	A	A	G	A	T	T	T	T	T	T	T	T	T	A	T	T	C	G	G	G	T	A	T
D19 Sequenza NCBI (13)		C A G A G A G A T T T T A T T C G G G T A A T G G G T G C A C C A A A T A T C A G A A A T C A C																								

Per valutare la precisione del metodo, si è valutata il valore medio e l'errore della misurazione dei tempi di ritenzione di analisi indipendenti. La precisione è indice della distribuzione di questi valori e permette, in maniera riflessa, di dare una valutazione del metodo nel senso di esprimere quanto questo sia preciso nelle chiamate dei picchi. Per fare ciò ho misurato cinque volte i tempi di ritenzione del gruppo EME da sangue umano spegnendo e riaccendendo lo strumento, e ricondizionando la colonna, in giorni differenti, così da valutare la stabilità del metodo (Figura 26).

Media Gruppo EME	12,256
Deviazione Standard	0,430
Intervallo 95% (2 σ)	0,861

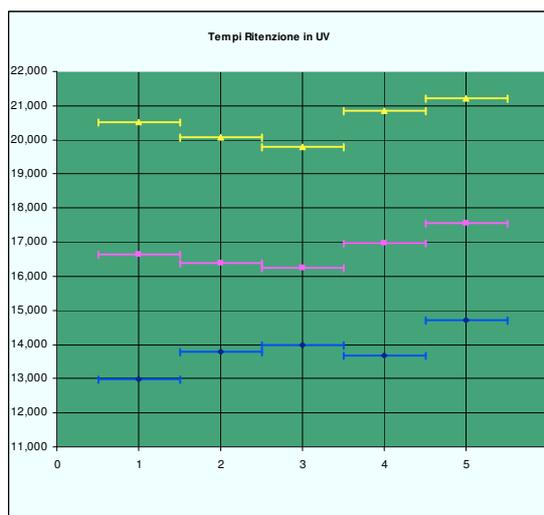


Figura 26: Sensibilità e precisione.

Dove i tratti blu sono i gruppi EME, quelli rosa sono la globulina β e quelli gialli sono la globulina α .

Valutando i tempi di ritenzione dei diversi picchi appare chiaro come con l'errore considerato i picchi siano ben distinguibili, e pertanto il metodo appare sufficientemente preciso.

L'accuratezza della metodica è stata valutando analizzando diversi animali della stessa specie anche di razze diverse. Si sono valutate una decina di animali per l'Oca, l'Anatra ed il Bovino, poi per i cani e gatti rispettivamente cinquantanove (Tabella 2) e ventiquattro di diverse specie. Nella Tabella 2 sono elencati per esempio i dati relativi ai 59 cani analizzati, mentre i risultati finali sono evidenziati nella Tabella 4. A fianco si è effettuato il rapporto con il tempo di ritenzione dell'EME, dato che per questo gruppo prostetico dovrebbe essere costante per tutte le specie, permettendo così di comparare meglio i risultati. Per tutte le specie considerate si sono ottenuti spettri simili, tranne per il gatto, per il quale invece la razza Europea presenta in alcuni casi un picco in più oltre alle classiche tre globuline. Questo risultato merita sicuramente un ulteriore approfondimento bio-molecolare.

Tabella 2: Tempi di ritenzione dei picchi rilevati in UV ed in Fluorescenza per 59 cani.

		EME	1° Picco	2° Picco	3° Picco		1°/EME	2°/EME	3°/EME
1	cane 1	12,093	13,227	14,08	14,26		1,094	1,164	1,179
2	cane 2	12,053	13,247	14,107	14,26		1,099	1,170	1,183
3	cane 3	12,067	13,227	14,087	14,267		1,096	1,167	1,182
4	cane 4	12,013	13,267	14,153	14,347		1,104	1,178	1,194
5	cane 5	12,013	13,293	14,16	14,347		1,107	1,179	1,194
6	cane 6	12,013	13,273	14,133	14,453		1,105	1,176	1,203
7	cane 8	12,04	13,273	14,14	14,3		1,102	1,174	1,188
8	cane 9	12,053	13,26	14,113	14,267		1,100	1,171	1,184

9	cane 12	13,56	15,827	16,68	16,867
10	cane 13	13,667	15,753	16,587	16,913
11	cane 14	13,6	15,84	16,7	16,86
12	cane 15	13,547	15,707	16,7	17,007
13	cane 16	13,593	15,827	16,687	16,887
14	cane 17	13,52	15,8	16,667	16,833
15	cane 18	13,567	15,8	16,693	17,027
16	cane 19	13,6	15,82	16,673	16,827
17	cane 20	13,553	15,793	16,647	16,833
18	cane 21	13,407	15,78	16,647	16,833
19	cane 22	13,607	15,753	16,607	16,94
20	cane 23	13,647	15,78	16,633	16,953
21	cane 24	13,68	15,613	16,613	16,773
22	cane 25	13,68	15,76	16,62	16,8
23	cane 28	13,687	15,787	16,64	16,787
24	cane 29	13,64	15,753	16,607	16,94

1,167	1,230	1,244
1,153	1,214	1,238
1,165	1,228	1,240
1,159	1,233	1,255
1,164	1,228	1,242
1,169	1,233	1,245
1,165	1,230	1,255
1,163	1,226	1,237
1,165	1,228	1,242
1,177	1,242	1,256
1,158	1,220	1,245
1,156	1,219	1,242
1,141	1,214	1,226
1,152	1,215	1,228
1,153	1,216	1,226
1,155	1,218	1,242

25	cane 33	13,653	15,733	16,587	16,9
26	cane 34	13,76	15,7	16,547	16,76
27	cane 35	13,727	15,733	16,62	16,76
28	cane 36	13,733	15,787	16,573	16,88
29	cane 37	13,793	15,753	16,573	16,85
30	cane 40	13,773	15,747	16,56	16,75
31	cane 41	13,767	15,833	16,66	16,9
32	cane 45	13,72	15,76	16,607	16,733
33	cane 47	13,653	15,68	16,527	16,68
34	cane 48	13,7	15,62	16,6	16,787
35	cane 50	13,613	15,787	16,647	16,973
36	cane 51	13,6	15,807	16,687	16,893
37	cane 53	13,533	15,807	16,667	16,86
38	cane 54	13,533	15,893	16,753	16,927
39	cane 55	13,313	16,007	16,927	17,16
40	cane 56	13,207	16,033	16,933	17,187

1,152	1,215	1,238
1,141	1,203	1,218
1,146	1,211	1,221
1,150	1,207	1,229
1,142	1,202	1,222
1,143	1,202	1,216
1,150	1,210	1,228
1,149	1,210	1,220
1,148	1,211	1,222
1,140	1,212	1,225
1,160	1,223	1,247
1,162	1,227	1,242
1,168	1,232	1,246
1,174	1,238	1,251
1,202	1,271	1,289
1,214	1,282	1,301

41	cane 57	13,187	16,087	17,018	17,387
42	cane 58	13,1	16,087	17,033	17,26
43	cane 59	13,127	16,107	17,053	17,273
44	cane 60	13,113	16,1	17,04	17,267
45	cane 61	13,1	16,09	17,067	17,3
46	cane 62	13,107	16,073	17,02	17,253
47	cane 65	13,167	16,12	17,067	17,3
48	cane 3-1	13,38	15,86	16,76	17,1
49	cane 3-2	13,34	15,887	16,787	16,973
50	cane 3-3	13,4	15,733	16,773	17,112
51	cane 3-6	13,307	15,907	16,807	17,033
52	cane 3-8	13,26	15,94	16,855	17,213
53	cane 3-9	13,293	15,907	16,813	17,02
54	cane 3-10	13,24	15,96	16,873	17,073
55	cane 3-11	13,227	15,953	16,86	17,093
56	cane 3-14	13,233	15,987	16,887	17,24

1,220	1,291	1,318
1,228	1,300	1,318
1,227	1,299	1,316
1,228	1,299	1,317
1,228	1,303	1,321
1,226	1,299	1,316
1,224	1,296	1,314
1,185	1,253	1,278
1,191	1,258	1,272
1,174	1,252	1,277
1,195	1,263	1,280
1,202	1,271	1,298
1,197	1,265	1,280
1,205	1,274	1,290
1,206	1,275	1,292
1,208	1,276	1,303

57	cane 3-15	13,247	15,98	16,907	17,147		1,206	1,276	1,294
58	cane 3-16	13,213	15,84	16,803	17,26		1,199	1,272	1,306
59	cane 3-17	13,28	15,96	16,807	17,067		1,202	1,266	1,285

	Media	13,276	15,504	16,391	16,626	Media	1,167	1,234	1,252
	Dev. St.	0,534	0,905	0,9184	0,9403	Dev. St.	0,038	0,039	0,040

La Sensibilità è il limite minimo di misurazione che permette di avere un risultato identificabile. Per fare ciò si sono effettuate delle diluizioni seriali di sangue umano che sono state analizzate con il rilevatore UV e con quello fluorimetrico (Figura 27 e 28).

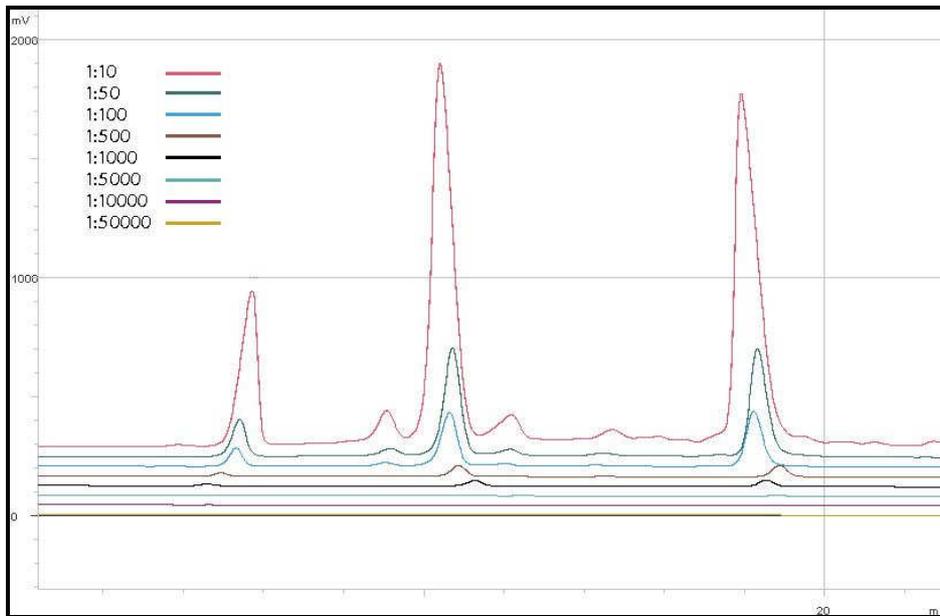


Figura 27: Scala contrazioni in UV, sangue umano fresco

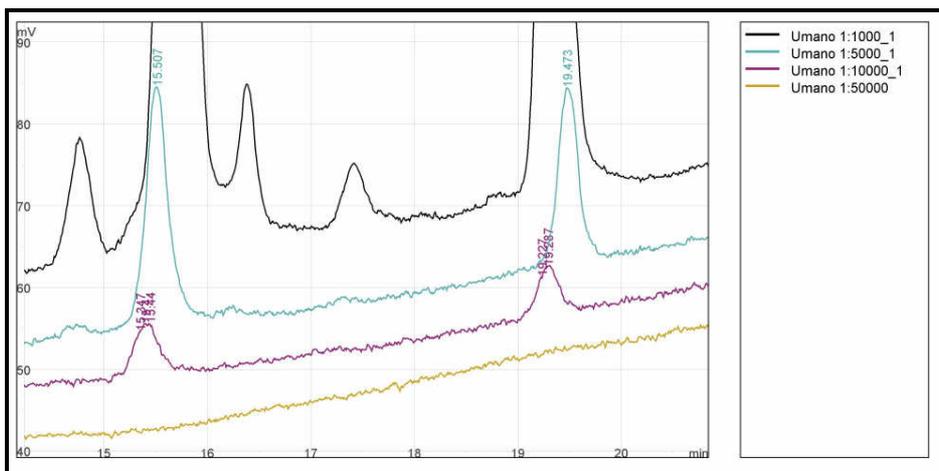


Figura 28: Sangue umano fresco alla concentrazione 1:10 in Fluorescenza

Nel lavoro di Johnston et. al (Comparison of Presumptive Blood Test Kits Including Hexagon OBTI) è stata effettuata una comparazione tra diversi kit per l'analisi delle tracce ematiche; il risultato mostra che l'OBTI è il kit più sensibile, con un limite a 1/10.000. L'analisi in DHPLC è perciò sicuramente paragonabile ai metodi immuno-cromatografici, riconosciuti generalmente essere i più sensibili (Tabella 3).

TABLE 1—Comparison of the sensitivity of four presumptive blood test kits on 50 μ L of dilute blood.

Blood Dilution	KM	LMG	Hemastix®	Hexagon OBTI
Neat	+	+	+	-
1 in 10	+	+	+	-
1 in 50	+	+	+	+
1 in 100	+	+	+	+
1 in 250	+	+	+	+
1 in 1×10^3	+	+	+	+
1 in 5×10^3	+	+	+	+*
1 in 1×10^4	+	+	+	+*
1 in 5×10^4	-	-	+	-
1 in 1×10^5	-	-	-	-

KM, Kastle-Meyer; LMG, Leucomalachite green; +, positive result; -, negative result.
 *Result after reducing the volume of buffer in collection bottle from 2 mL to 200 μ L.

Tabella 3: Comparazione tra diversi Kit per l'analisi delle tracce ematiche.

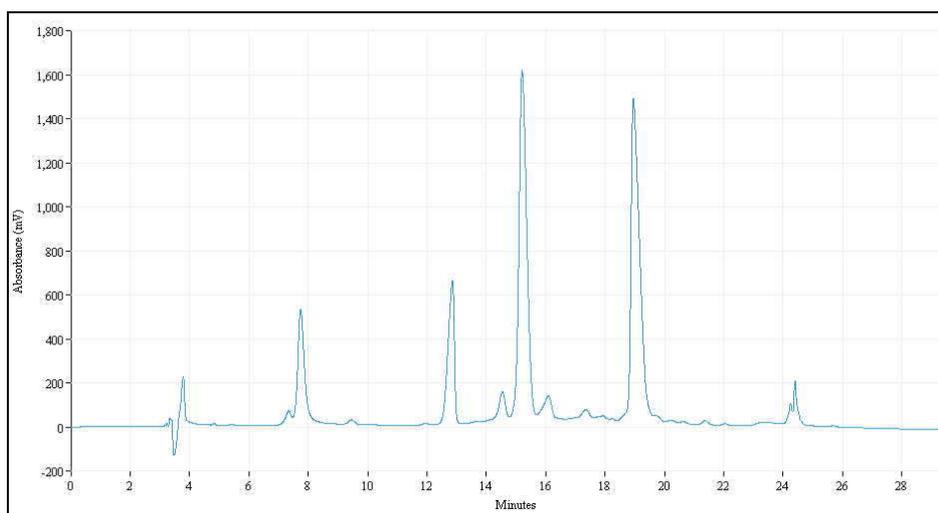


Figura 29: Sangue umano fresco alla concentrazione 1:10 in UV.

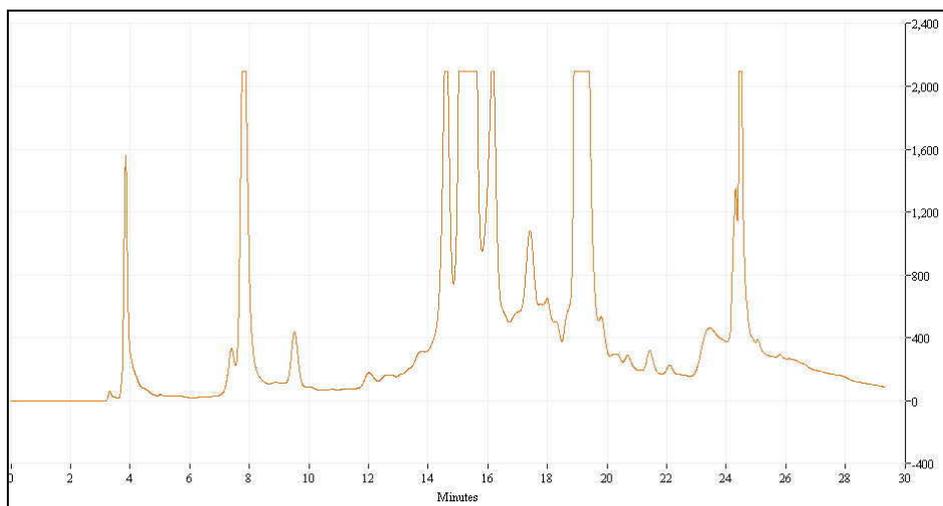


Figura 30: Sangue umano fresco alla concentrazione 1:10 in FL.

Per specificità si intende la capacità di identificare correttamente quello che si misura. Per fare ciò si è creato un pattern di emoglobine, scegliendo tra gli animali che sono maggiormente diffusi nell'ambiente domestico, e che quindi potrebbero con maggiore probabilità lasciare delle tracce ematiche nel luogo del reato. Tra gli animali più frequenti citiamo: Ratto, Cavia, Cane, Gatto, Bovino, Suino, Anatra, Oca.

I campioni di sangue fresco sono stati analizzati ad una concentrazione di 1:100 e di 1:300 e sono stati confrontati con il pattern umano delle globine, per verificare se l'analisi fosse effettivamente specie-specifica. Ancora una volta oltre alle analisi in UV e FL si sono effettuati i rapporti per i tempi di ritenzione dei gruppi EME di dette corse.

Dall'analisi in UV, che permette anche l'individuazione dell'Eme presente nell'emoglobina (Figura 31), sono stati estrapolati i valori in tabella 4.

ANALISI IN UV (220 nm)	picco - 1	picco - 2	EME	1°picco	2°picco	3°picco
Umano			12,3	15,2	17,4	
Ratto	10,6	11,9	12,1	12,2	13	13,2
Cavia			12,1	14,9	15	16,6
Cane			12,2	13,2	14,1	14,2
Gatto			12,1	13,6	15	16,5
Bovino			13	18,1	21,2	
Suino			11,9	13,9	18,2	
Anatra			11,7	13,8	17,2	
Oca			12,9	17,6	18,6	

Tabella 4: Media dei tempi di ritenzione.

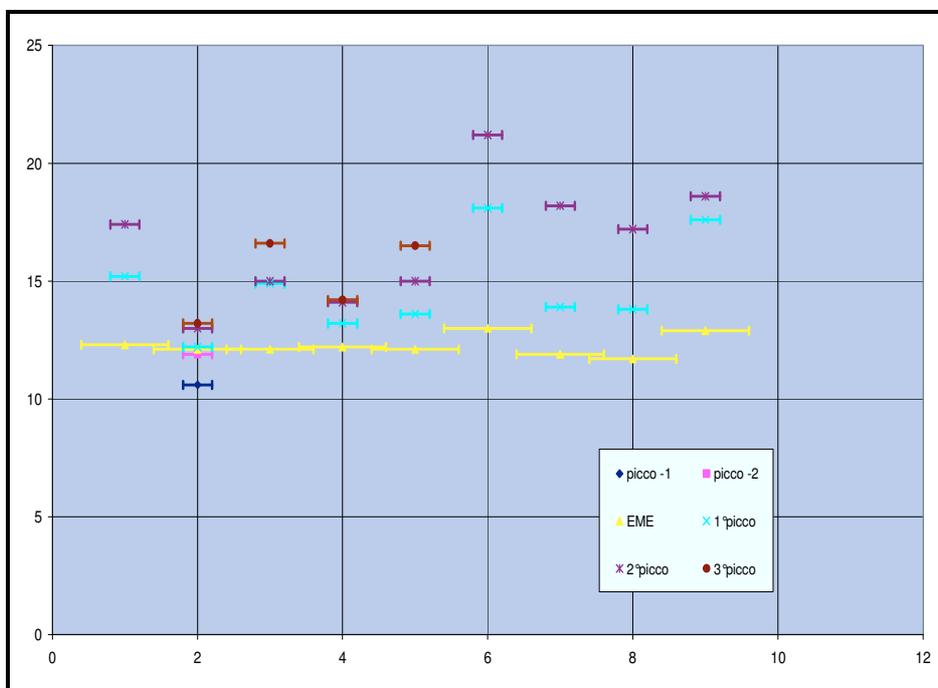


Figura 31: Grafico in sequenza delle medie dei tempi di ritenzione delle diverse specie animali con rivelazione in UV.

I valori sono poi stati normalizzati (Figura 32 e Tabella 5) per essere standardizzati e per poterli confrontare al meglio, evitando così di avere differenze nei tempi di ritenzione delle globine. La stessa cosa è stata poi effettuata anche in Fluorescenza (Figura 33, 34 e Tabella 6, 7), dove la normalizzazione è stata effettuata dividendo per il tempo di ritenzione del gruppo EME della medesima corsa recuperata dall'analisi in UV.

ANALISI IN UV (220 nm)	-1°/eme	-2°/eme	EME	1°/eme	2°/eme	3°/eme
Umano			1	1,235	1,414	
Ratto	0,876	0,98	1	1,008	1,074	1,909
Cavia			1	1,231	1,239	1,371
Cane			1	1,081	1,155	1,163
Gatto			1	1,123	1,239	1,363
Bovino			1	1,392	1,63	
Suino			1	1,168	1,529	
Anatra			1	1,179	1,47	
Oca			1	1,364	1,441	

Tabella 5: Media dei tempi di ritenzione normalizzata per l'EME in UV.

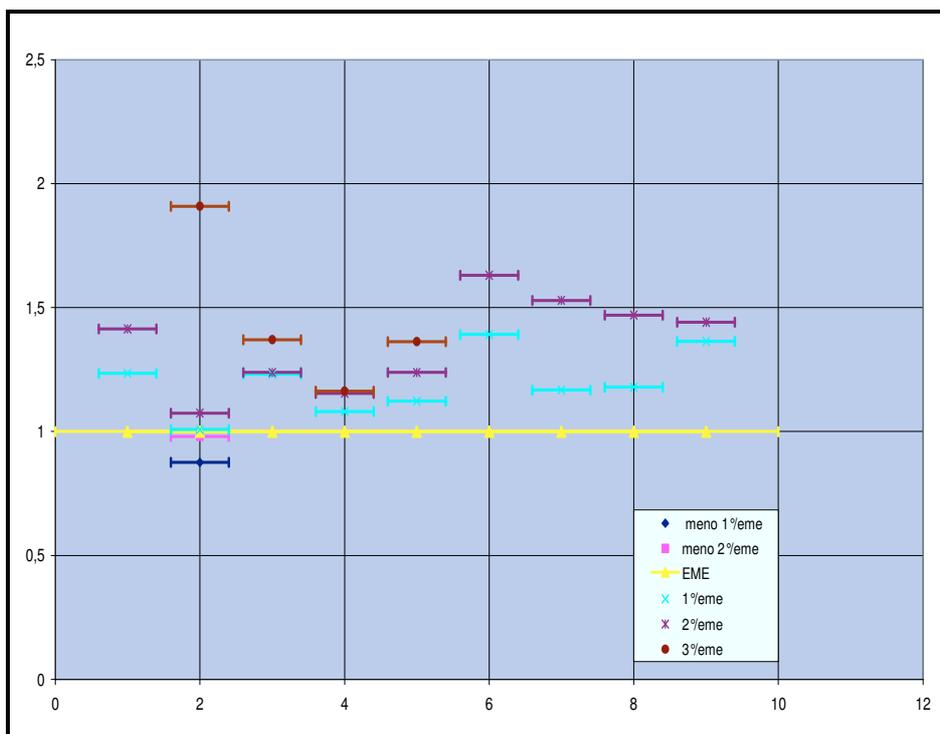


Figura 32: Grafico della normalizzazione in sequenza dei tempi di ritenzione delle diverse specie animali con rivelazione in UV.

ANALISI IN FLUORESCENZA (275/345 nm)	picco -1	picco -2	EME	1°picco	2°picco	3°picco
Umano				15,5	18	
Ratto	10,9	11,7		12,3	13	13,1
Cavia				14,9	15	16,6
Cane				13,3	14,1	14,3
Gatto				13,7	15,1	16,6
Bovino				18,1	21,4	
Suino				12,8	18,1	
Anatra				13,9	17,4	
Oca				17,7	18,4	

Tabella 6: Medie dei tempi di ritenzione normalizzata per l'EME in Fluorescenza.

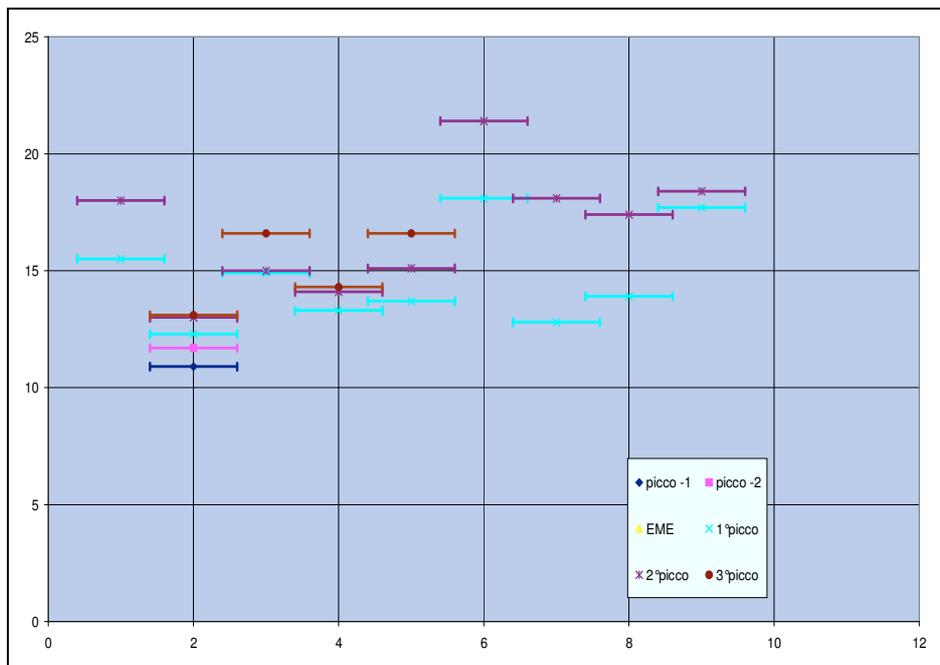


Figura 33: Grafico dei tempi di ritenzione in sequenza delle diverse specie animali con rivelazione in Fluorescenza.

ANALISI IN FLUORESCENZA (275/345 nm)	EME (basandosi sulla media in UV di 12,2)			1°picco	2°picco	3°picco
	picco -1	picco -2				
Umano			1	1,27	1,475	
Ratto	0,893	0,959	1	1,008	1,065	1,073
Cavia			1	1,221	1,229	1,36
Cane			1	1,09	1,55	1,172
Gatto			1	1,22	1,237	1,36
Bovino			1	1,483	1,754	
Suino			1	1,049	1,483	
Anatra			1	1,139	1,426	
Oca			1	1,45	1,508	

Tabella 7: Medie dei tempi di ritenzione normalizzata per l'EME in Fluorescenza.

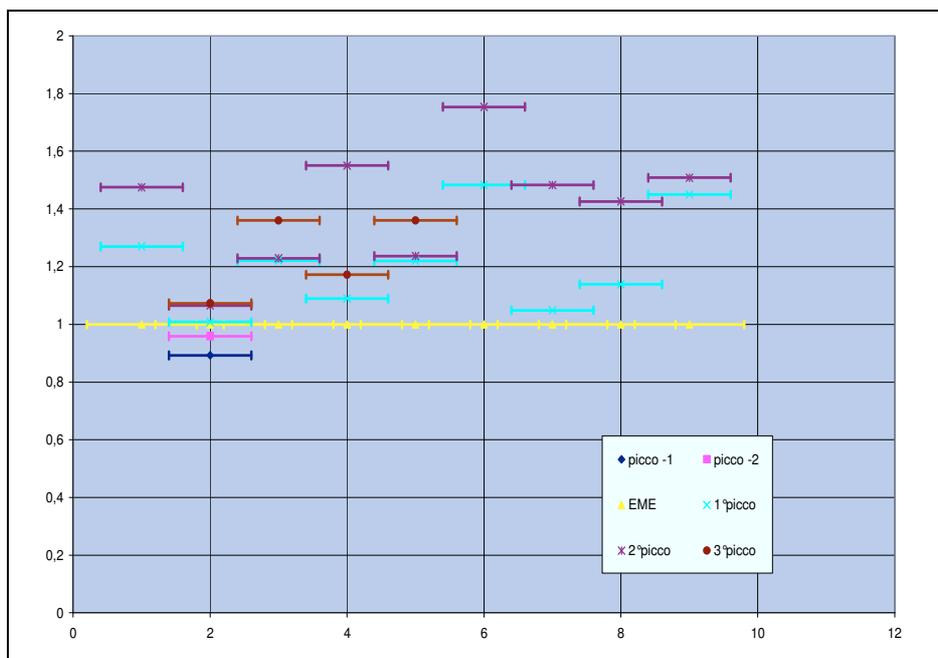


Figura 34: Normalizzazione delle analisi in sequenza dei campioni delle diverse specie animali con rivelazione in Fluorescenza

Come si può notare, la corsa dell'emoglobina della specie umana è nettamente distinguibile dalle altre specie, ma anche le altre specie tra loro possono essere facilmente identificate. È chiaro che una volta analizzata una specie ignota e averla confrontata con questo pattern, se si dovesse riscontrare una possibile similitudine, si deve correre detta specie nota per ulteriore conferma (EURACHEM Guide (1998) The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics, p. 42).

I campioni con Combur positivo e OBTI negativo di alcuni casi reali sono stati analizzati tramite DHPLC per verificare l'effettiva presenza di sangue e l'eventuale caratterizzazione della specie.

I campioni prelevati sono stati immersi in acqua distillata (dai 50 ai 100 µL) e vortexati; dopo un riposo di 30 minuti a 37°C, sono stati centrifugati a 14.000 rcf per 5 minuti. Il surnatante è stato prelevato e iniettato completamente nel DHPLC.

Nel primo caso, il campione sembrava essere sangue coagulato su un cartone, trovato accanto a una testa di maiale usata a scopo intimidatorio; il Combur risultava positivo, mentre l'OBTI risultava negativo. Dall'analisi in DHPLC si è potuto appurare che effettivamente si trattava di sangue, e dal

confronto con la tabella creata con i diversi tempi di ritenzione, si è potuto dire che si trattava di sangue di maiale (Figura 35, 36, 37 e 38).

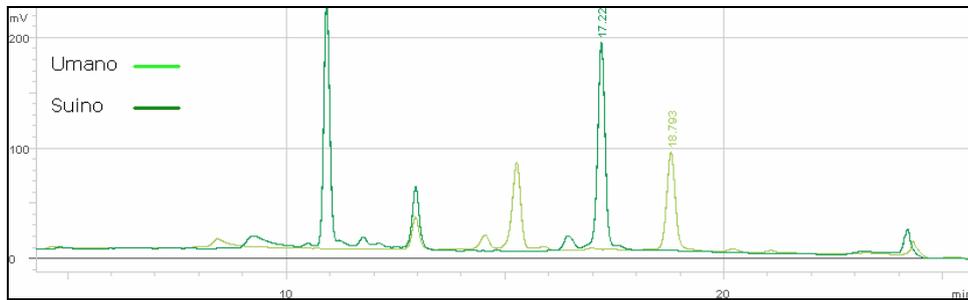


Figura 35: Sangue umano e sangue suino in UV

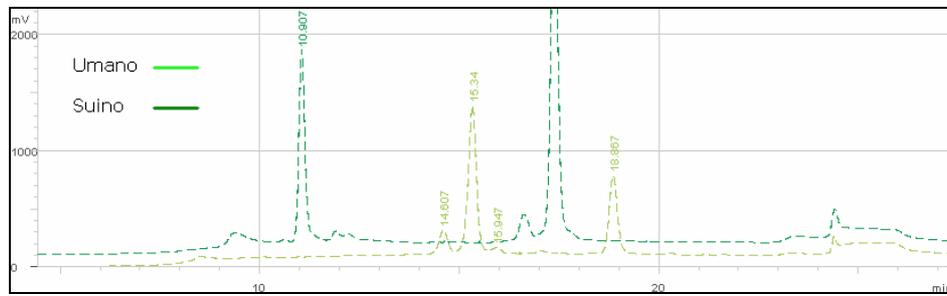


Figura 36: Sangue umano e sangue suino in Fluorescenza

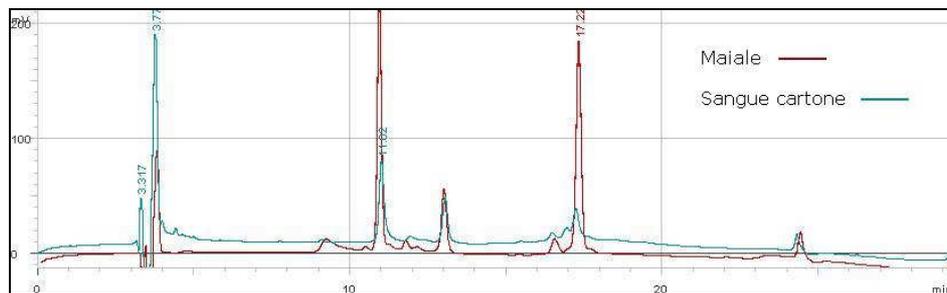


Figura 37: Sangue suino e tracce ematiche prelevate dal campione in UV

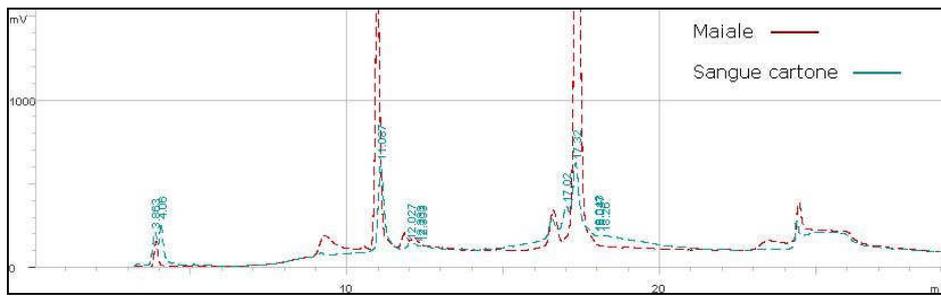


Figura 38: Sangue suino e tracce ematiche prelevate dal campione in Fluorescenza

Nel secondo caso, il campione è stato prelevato da una giacca e da un cappotto di un partigiano. Il Combur risultava positivo, l'OBTI negativo, probabilmente a causa della scarsa quantità di materiale ematico presente e della degradazione di questo essendo tracce vecchie di più di sessanta anni; l'analisi in DHPLC ha dato infatti esito positivo per entrambi i campioni, confermando perciò che fosse sangue, e da una valutazione della posizione dei fori si è potuto stabilire che fu un'esecuzione invece che uno scontro a fuoco, chiarendo così i fatti dopo moltissimi anni (Figura 39, 40, 41 e 42).

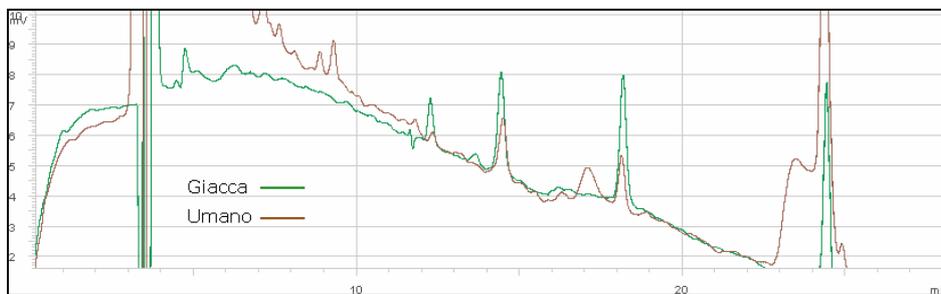


Figura 39: Sangue umano e tracce ematiche prelevate dal primo campione in UV.

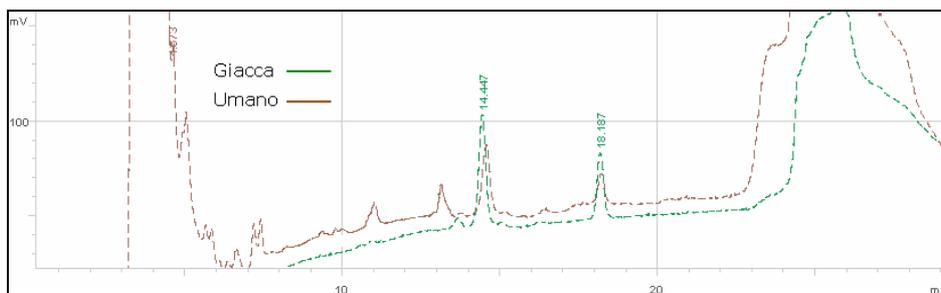


Figura 40: Sangue umano e tracce ematiche prelevate dal primo campione in Fluorescenza.

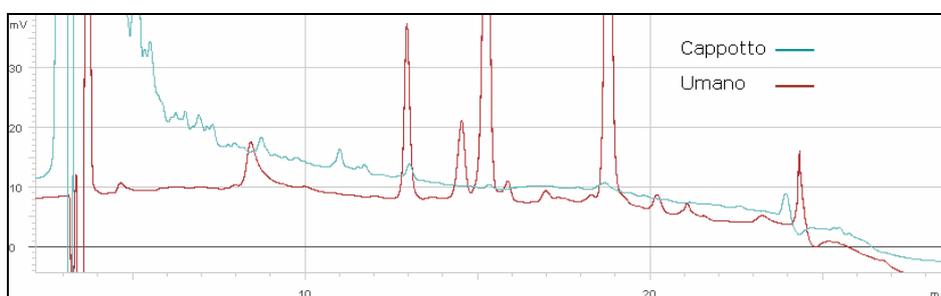


Figura 41: Sangue umano e tracce ematiche prelevate dal secondo campione in UV.

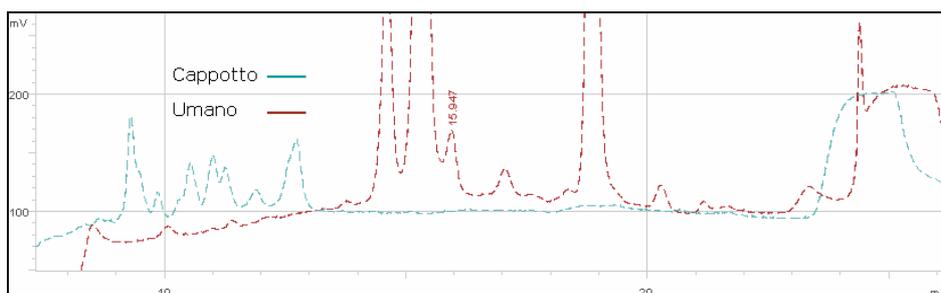


Figura 42: Sangue umano e tracce ematiche prelevate dal secondo campione in Fluorescenza.

DISCUSSIONE

Nel mio studio ho indagato nel genoma umano le microvariante alleliche rare del marcatore STR D19S433 da un campione di popolazione umana di circa 30.000 soggetti. Per questo scopo ho analizzato nella popolazione umana 30.000 campioni di DNA. La microvariante 6.2 da me indagata è già nota, come descritto nel sito del NIST o National Institute of Standard and Technologies nella parte dedicata agli STR: http://www.cstl.nist.gov/strbase/var_tab.htm). La sua frequenza, riportata dal sito del NIST (National Institute of Standard and Technology), è di circa 1/11.500 persone prese a caso nella popolazione mondiale. Nel mio campione di popolazione ho osservato che tale microvariante mantiene la stessa frequenza. Questo indica che il campione da me preso in esame è numericamente sufficiente per costituire un campione statisticamente significativo. I quattro diversi alleli 6.2 sequenziati sono comunque un numero abbastanza piccolo, ma sufficiente, per avere una valutazione del tipo di mutazioni osservate. Le sequenze di quattro alleli come indicato nei Risultati sono tra loro identiche. Queste sequenze sono inoltre identiche a quelle identificate da Heinrich e collaboratori (2005). Heinrich et al., verificarono che i quattro soggetti a cui appartenevano le sequenze indagate erano tutti di origine Turca. I campioni biologici impiegati nel mio studio non sono catalogati per origine etnica. Pertanto non mi è possibile fare speculazioni su un'eventuale area di origine comune tra gli individui studiati da me e da Heinrich et al. Questi stessi autori, in uno studio effettuato sulla popolazione Turca, non riscontrano l'allele off-ladders (microvariante). Però, il campione indagato non era statisticamente significativo per il numero degli individui analizzati. Pertanto, questo dato va considerato principalmente alla luce della scarsa frequenza con cui questa micro variante è presente nella popolazione mondiale. Sulla scorta di queste considerazioni ipotizzo che la variante da me analizzata sia da considerarsi come una mutazione, piuttosto che un polimorfismo. Infatti, è probabile che la mutazione una volta fissata sul genoma nella sequenza considerata si presenti come una delezione che avviene in modo preferenziale poche basi prima della sequenza ripetuta. Ritengo plausibile che la mutazione, come quella che causa variazioni nella lunghezza del motivo ripetuto, avvenga per il fenomeno di slippage durante la fase di duplicazione del DNA. Infatti, mutazioni di questo tipo sono state identificate a poche paia di basi dall'inizio della sequenza ripetuta. Il fenomeno che causa queste variazioni di lunghezza provoca, dunque, livelli

abbastanza alti di polimorfismo per lunghezza come riportato nelle banche dati internazionali (CODIS). Tale caratteristica rende i microsatelliti dei buoni marcatori molecolari codominanti negli studi di genetica di popolazioni. Inoltre, la semplice discriminazione tra alleli di dimensioni differenti tramite elettroforesi li rende facilmente analizzabili. Il processo definito “Slipped-Strand Mispairing” (SSM) consiste in una denaturazione locale della doppia elica, un appaiamento scorretto dei filamenti con un slittamento di alcune basi e la successiva rinaturazione della doppia elica. Nel loop che si forma dal mancato appaiamento dei nucleotidi complementari, dovuto allo “slippage”, non avvengono riparazioni fino al successivo ciclo di replicazione. In tal modo si ha l’aggiunta di una ripetizione quando il loop si trova sul filamento di allungamento e viceversa una delezione della stessa unità ripetitiva se si trova sul filamento stampo. Più lunga è l’unità ripetitiva maggiormente distorta risulterà la doppia elica allineata scorrettamente. Wells e collaboratori (1989) hanno riscontrato che i tassi di inserzioni di unità ripetute diminuiscono quando il motivo ripetuto passa da 2 a 4 basi, in accordo con un principio di riduzione della distorsione nel DNA, per cui sono sfavoriti gli eventi che interessano distorsioni della doppia elica più lunghi. Pumpernik e collaboratori (2008) attraverso un confronto diretto delle differenze tra il genoma dello scimpanzè e quello umano hanno calcolato il grado di incidenza delle mutazioni puntiformi (transizioni, trasversioni, delezioni e inserzioni) contro lo slippage replicativo. Dal loro studio, sebbene focalizzato sugli STR mono, di e trinucleotidi risulta che lo slippage replicativo avviene più facilmente di un ordine di grandezza rispetto alle mutazioni puntiformi, tranne a basso numero di replicazioni (5-9 unità ripetute) dove sono dominanti le mutazioni puntiformi. Questo dato suggerisce che la microvariante 6.2 si sia generata per slippage replicativo da un allele più grande, come per esempio 14.2 o 15.2, che già presentava la delezione AG identificata.

In alternativa, prendendo in considerazione il numero standard dei polimorfismi più o meno frequenti (circa 20), la frequenza media di un allele è comunque bassa (circa 5%). Il risultato della mia ricerca mostra che tutte e quattro le sequenze, sebbene molto rare, sono identiche e come detto sono identiche a quelle identificate da Heinrich (2005). Questi aspetti, in associazione con il risultato sperimentale da me ottenuto, mi spingono a considerare, visto che la frequenza della microvariante 6.2 è di circa 0,01%, che il mio STR potrebbe anche essere un polimorfismo e come suggerito da Heinrich (2005) derivato da

popolazioni Medio Orientali. Ulteriori indagini sono quindi necessarie per accertare se siamo in presenza di una rara mutazione o di un raro polimorfismo.

Per quanto concerne le emoglobine, il metodo di studio da me impiegato è stato valutato per molti aspetti differenti, dando per tutti una risposta positiva. Si sono valutate le variazioni al passare del tempo. Si sono valutate le variabilità tra specie differenti e tra etnie diverse dell'uomo o di razze animali diverse per la stessa specie. Tra le tredici specie analizzate solo il gatto ha mostrato delle differenze, che comunque non rendevano dubbia l'identificabilità. Si è valutata poi la capacità di dare una risposta al diminuire della concentrazione dando una risposta ampiamente paragonabile a kit di identificazione della specie umana basati su tecniche di immunocromatografia, che sono sì sensibili ma con un costo di applicazione di almeno cinque volte superiore (circa 1 euro contro 5 euro del kit OBTI). Permette inoltre di non perdere il DNA, in quanto una volta solubilizzata la traccia, si può centrifugare (15 minuti a 15.000 g oppure 5 min a 25.000 a 4°C); recuperare il surnatante con le proteine disciolte e tenere il DNA sedimentato sul fondo per le eventuali analisi genetiche. E' stato inoltre impiegato in due casi reali, e anche in questi casi oltre ad identificare chiaramente la presenza del sangue, ha potuto darci anche una stima relativa alla specie di appartenenza della traccia ematica. Pertanto, ritengo che detto saggio abbia i requisiti minimi indispensabili per essere valutato come metodo identificativo della presenza di sangue e diagnostico della specie di appartenenza di detta traccia se il risultato è positivo. Ritengo quindi che lo studio da me eseguito sulle emoglobine con il D-HPLC, potrà avere una ricaduta applicativa in genetica forense.

RINGRAZIAMENTI

Per iniziare vorrei ringraziare le due persone che mi hanno permesso di raggiungere questa importante traguardo. Non solo per avermi dato la possibilità di poter lavorare nei loro laboratori ma anche per avermi guidato e stimolato per quella che io ritengo una crescita morale ancorché scientifica, il raggiungimento di un traguardo che però, da questo punto di vista, risulta più simile a dei blocchi di partenza. Queste persone, elencate così solo in ordine alfabetico, sono il Magg. Inv.sc. Marco Pizzamiglio, C.te della Sezione di Biologia del RIS Carabinieri di Parma, e il Prof. Mauro Tognon, Professore Ordinario di Biologia Applicata dell'Università degli Studi di Ferrara.

Ovviamente il più caro ringraziamento va ai miei famigliari mio padre Francesco, mia madre Paola, mia sorella Stefania e la mia fidanzata Federica, che ovviamente sono i primi ad aver sostenuto, ma soprattutto sopportato, questo percorso.

Poi i tesisti che, anche se per brevi periodi, hanno contribuito notevolmente, in modo particolare ringrazio Natalia Lugli con la sua costante energia, i buoni propositi e competenza, e Cornetti Luca dottorando, ora Dottore, dell'Università degli Studi di Ferrara.

Ci sono altri poi che hanno contribuito, chi più chi meno, che cercherò di elencare in ordine casuale certo di dimenticare qualcuno: Gen.B. Luciano Garofano, Ex C.te RIS Carabinieri di Parma per la sua guida illuminante, il Dr. Davide Grisaffi per avermi mille volte sospinto, la famiglia Vanni Reggiani, Arianna e Valentina che ormai mi hanno adottato, e tutti i colleghi, passati e presenti, della Sezione di Biologia del RIS Carabinieri di Parma con i vari gradi non citati, elencati per nome: Alberto, Alessandro, Fabiano, Giorgio, Nicola, Tiziano, Claudio, Massimiliano, Valentino, Davide, Paolo, Giancarlo, Diego, Simone, Giovanni, Domenico, Gianluca, Dario e Carmine; i quali più di tutti mi devono sopportare per molte ore tutti i giorni. A tutti Voi un grazie di cuore!

BIBLIOGRAFIA

Andrasko J., Rosén B. 1994. Sensitive identification of haemoglobin in bloodstains from different species by high performance liquid chromatography with combined UV and fluorescence detection. *J Forensic Sci.* 39(4):1018-25.

Bassam J.S., Caetano-Anollés G., Gresshoff P.M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem.* 196:80-3.

Beckman J.S., Weber J.L. 1992. Survey of human and rat microsatellites. *Genomics.* 12:627-31.

Brinkmann B., Klintschar M., Franz Neuhuber F., Hühne J., Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet.* 62:1408-15.

Deka R.D., Shriver M.D., Yu L.M., Ferrell R.E., Chakraborty R. 1995. Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis.* 16:1659-64.

Di Rienzo A., Donnelly P., Toomajian C., Sisk B., Adrian Hill A., Petzl-Erler M.L., Haines G.K., Barch D.H. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics.* 148:1269-84.

Doris P.A., Oefner P.J., Chilton B.S., Hayward-Lester A. 1998. Quantitative analysis of gene expression by ion-pair high-performance liquid chromatography. *J Chromatogr A.* 806:47-60.

Economou E.P., Bergen A.W., Warren A.C., Antonarakis S.E. 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA.* 87:2951-4.

- Ellis N., Goodfellow P.N.** 1989a. The mammalian pseudoautosomal region. *Trends Genet.* 5:406-410.
- Freije D., Helms C., Watson M.S., Donis-Keller H.** 1992. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Sci.* 258:1784-1787.
- Heinrich M., Felske-Zech H., Brinkmann B., Hohoff C.** 2005. Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. *Int J Legal Med.* 119:310-313.
- Heyer E., Puymirat J., Dieltjes P., Bakker E., De Knijff P.** 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet.* 6:799-803.
- Hoogendoorn B., Owen M.J., Oefner P.J., Williams N., Austin J., O'Donovan M.C.** 1999. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Hum Genet.* 104:89-93.
- Hoskins R.A., Phan A.C., Naemuddin M., Mapa F.A., Ruddy D.A., Ryan J.J., Young L.M., Wells T., Koczynski C., Ellis M.C.** 2001. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* 11:1100-13.
- Litt M., Luty J.A.** 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet.* 44:397-401.
- Morton N.E.** 1991. Parameters of the human genome. *Proc Natl Acad Sci USA.* 88:7474-7476.
- Mullis K.B.** 1990. Target amplification for DNA analysis by the polymerase chain reaction. *Ann. Biol. Clin. Paris.* 48:579-82.

Premstaller A., Oefner P.J. 2002. Denaturing HPLC for nucleic acids. LC-GC Europe. p410, 10p.

Parson W., Pegoraro K., Niederstatter H., Foger M. 2000. Species identification by means of the cytochrome b gene. *Int J Legal Med.* 114:23-28.

Pumpernik D., Oblak B., Borstnik B. 2008. Replication slippage versus point mutation rates in short tandem repeats of the human genome. *Mol Genet Genomics.* 279:53-61.

Sambrook J., Russel M. 1989. *Molecular cloning. A laboratory manual, second ed., vol. I, II, III.* Cold Spring Harbor Laboratory Press, USA.

Schriml L.M., Dean M. 2000. Identification of 18 mouse ABC genes and characterization of the ABC superfamily in *Mus musculus*. *Genomics.* 64:24-31.

Spiegelman J.I., Mindrinos M.N., Fankhauser C., Richards D., Lutes J., Chory J., Oefner P.J. 2000. Cloning of the *Arabidopsis* RSF1 gene by using a mapping strategy based on high-density DNA arrays and denaturing high-performance liquid chromatography. *Plant Cell.* 12:2485-2498.

Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J., McCusker J.H., Davis R.W. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature.* 416:326-30.

Stoneking M. 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet.* 67:1029-1032.

Vosberg H.P. 1989. The polymerase chain reaction: An improved method for the analysis of nucleic acids. *Hum Genet.* 83:1-15.

Weber J., May P. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet.* 44:388-96.

Webster M.T., Smith N.G., Ellegren H. 2002. Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. *Proc Natl Acad Sci U S A.* 99:8748-53.

Wells R.A., Green P., Reeders S.T. 1989. Simultaneous genetic mapping of multiple human minisatellite sequences using DNA fingerprinting. *Genomics.* 5(4):761-72.

Xiao W., Oefner P.J. 2001. Denaturing high-performance liquid chromatography: A review. *Hum Mutat.* 17:439-74.

Xu X., Peng M., Fang Z. 2000. The direction of microsatellite mutations is dependent upon allele length. *Nat Genet.* 24:396-9.