



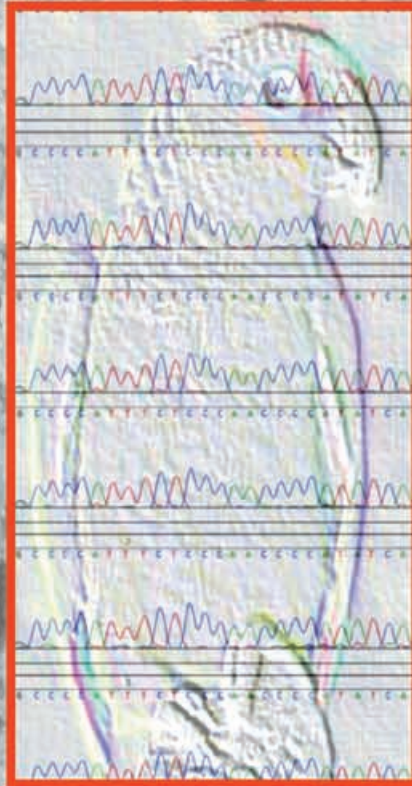
MINISTERO DELL'AMBIENTE
E DELLA TUTELA DEL TERRITORIO
Direzione Conservazione della Natura



ISTITUTO NAZIONALE
PER LA FAUNA SELVATICA
"ALESSANDRO GHIGI"

Ettore Randi, Cristiano Tabarroni e Silvia Rimondi

Genetica forense in applicazione della Convenzione di Washington CITES



Quaderni di Conservazione della Natura

La collana “Quaderni di Conservazione della Natura” nasce dalla collaborazione instaurata tra il Ministero dell’Ambiente, Servizio Conservazione della Natura e l’Istituto Nazionale per la Fauna Selvatica “A. Ghigi”. Scopo della collana è quello di divulgare le strategie di tutela e gestione del patrimonio faunistico nazionale elaborate dal Ministero con il contributo scientifico e tecnico dell’I.N.F.S.

I temi trattati spaziano da quelli di carattere generale, che seguono un approccio multidisciplinare ed il più possibile olistico, a quelli dedicati a problemi specifici di gestione o alla conservazione di singole specie.

This publication series, specifically focused on conservation problems of Italian wildlife, is the result of a co-operation between the Nature Conservation Service of the Italian Ministry of Environment and the National Wildlife Institute “A. Ghigi”. Aim of the series is to promote a wide circulation of the strategies for the wildlife preservation and management worked up by the Ministry of Environment with the scientific and technical support of the National Wildlife Institute.

The issues covered by this series range from general aspects, based on a multidisciplinary and holistic approach, to management and conservation problems at specific level.

COMITATO EDITORIALE

ALDO COSENTINO, ALESSANDRO LA POSTA, MARIO SPAGNESI, SILVANO TOSO

In copertina: Elaborazione grafica di Cristiano Tabarroni.

MINISTERO DELL'AMBIENTE
E DELLA TUTELA DEL TERRITORIO
DIREZIONE CONSERVAZIONE DELLA NATURA

ISTITUTO NAZIONALE PER LA
FAUNA SELVATICA "A. GHIGI"

Ettore Randi, Cristiano Tabarroni e Silvia Rimondi

Genetica forense in applicazione
della Convenzione di Washington CITES

QUADERNI DI CONSERVAZIONE DELLA NATURA
NUMERO 12

La redazione raccomanda per le citazioni di questo volume la seguente dizione:

Randi E., C. Tabarroni e S. Rimondi, 2002 - *Genetica forense in applicazione della Convenzione di Washington CITES*. Quad. Cons. Natura, 12, Min. Ambiente - Ist. Naz. Fauna Selvatica.

Tutti i diritti sono riservati. Nessuna parte di questa pubblicazione può essere riprodotta, memorizzata o trasmessa con qualsiasi mezzo e in qualsiasi forma (elettronica, elettrica, chimica, meccanica, ottica, fotostatica) o in altro modo senza la preventiva autorizzazione del Ministero dell'Ambiente e della Tutela del Territorio.

Vietata la vendita: pubblicazione distribuita gratuitamente dal Ministero dell'Ambiente e della Tutela del Territorio e dall'Istituto Nazionale per la Fauna Selvatica "A. Ghigi".

INDICE

INTRODUZIONE	Pag.	7
LA CONVENZIONE SUL COMMERCIO INTERNAZIONALE DELLE SPECIE ANIMALI E VEGETALI SELVATICHE MINACCIATE D'ESTINZIONE (CITES)	"	8
La CITES	"	8
STRUTTURA E FUNZIONI DEL DNA	"	18
Genetica forense e DNA fingerprinting	"	18
Introduzione al DNA fingerprinting	"	20
Struttura e funzioni del DNA	"	22
Mutazioni e polimorfismi genetici	"	39
VARIABILITÀ GENETICA NEGLI INDIVIDUI E NELLE POPOLAZIONI	"	41
I meccanismi dell'eredità: le leggi di Mendel	"	41
I meccanismi dell'eredità: associazione fra geni	"	45
I geni nelle popolazioni	"	48
Inbreeding	"	51
GENETICA MOLECOLARE: METODI DI ANALISI DELLA VARIABILITÀ GENETICA A LIVELLO DEL DNA	"	57
Metodi di raccolta dei campioni biologici	"	57
Metodi di raccolta delle tracce	"	61
Conservazione dei campioni	"	62
Estrazione del DNA	"	65
Controllo dell'estrazione del DNA	"	71
Enzimi di restrizione ed analisi dei polimorfismi dei frammenti di restrizione	"	71
Analisi dei frammenti di DNA tramite elettroforesi su gel di agarosio	"	74
Southern blotting	"	77
Ibridazione molecolare	"	77
Il protocollo INFS per le analisi di DNA fingerprinting multilocus	"	79
Struttura delle sonde multilocus usate in genetica forense ..	"	83
Interpretazione dei DNA fingerprinting	"	85
Amplificazione del DNA	"	87

RAPD (Random Amplified Polymorphic DNA)	Pag. 91
AFLP (Amplified Fragment Length Polymorphism)	" 92
Sequenziamento del DNA	" 92
Struttura e sequenziamento del DNA mitocondriale	" 98
Amplificazione ed analisi dei microsatelliti	" 99
Analisi dei microsatelliti nei sequenziatori automatici	" 100
Cromosomi sessuali e determinazione genetica del sesso	" 103
Determinazione della similarità fra frammenti di DNA, alleli, genotipi (aplotipi e genotipi multilocus) ed individui	" 107
Identificazione dei frammenti di DNA in analisi di DNA fingerprinting con sonde multilocus	" 107
Stima delle frequenze alleliche tramite binning in sistemi multilocus	" 108
Identificazione degli alleli in analisi di DNA fingerprinting con sistemi VNTR	" 110
Identificazione degli alleli in analisi di DNA fingerprinting con microsatelliti (STR)	" 110
ANALISI STATISTICA DEI DATI	" 112
Distribuzioni di frequenza	" 112
Stima degli intervalli di confidenza	" 116
Test di significatività	" 117
Stima delle frequenze alleliche e genotipiche	" 118
Stima delle frequenze alleliche a loci codominanti (VNTR o STR)	" 121
Stima delle frequenze alleliche in minisatelliti analizzati con sistemi MLP	" 121
Stima delle frequenze genotipiche a sistemi multilocus	" 122
PROBABILITÀ	" 124
Teoria frequentistica della probabilità	" 124
Teoria soggettivistica della probabilità (statistica Bayesiana)	" 125
Le regole del calcolo delle probabilità	" 125
Il teorema di Bayes	" 127
APPLICAZIONI DELLA STATISTICA BAYESIANA ALLA GENETICA FORENSE	" 129
Identificazione	" 129
Probabilità di esclusione	" 133
Probabilità di match	" 135
Probabilità di identità	" 138
Test di paternità	" 139

CASI DI STUDIO	Pag. 143
Test di parentela effettuati tramite DNA fingerprinting	" 143
Test di parentela effettuati tramite microsatelliti	" 146
Identificazione di sottospecie di scimpanzè mediante analisi del mtDNA.	" 146
EXECUTIVE SUMMARY	" 149
BIBLIOGRAFIA	" 151

INTRODUZIONE

La genetica forense sta attraversando un periodo di rapidi progressi grazie allo sviluppo delle metodiche d'analisi molecolare del DNA. Le analisi della variabilità genetica a livello del DNA hanno raggiunto gradi di precisione, ripetibilità ed affidabilità che non erano immaginabili fino a pochi anni fa. Il concetto di DNA fingerprinting, l'impronta digitale genetica, è entrato rapidamente a far parte del linguaggio comune. I punti di forza delle metodiche molecolari sono: l'elevata capacità d'individualizzazione (ogni individuo, eccetto i gemelli monozigoti, è dotato di un patrimonio genetico unico, che lo differenzia da qualsiasi altro individuo), e la possibilità di analizzare ed interpretare i risultati delle analisi di laboratorio nel contesto della genetica delle popolazioni e della teoria delle probabilità. In questo modo i risultati delle analisi di laboratorio possono essere espressi quantitativamente (probabilisticamente) e valutati tramite test statistici.

Il significato ed il valore del DNA fingerprinting sono stati a lungo dibattuti nella letteratura scientifica internazionale. Gli aspetti tecnici, riguardanti l'affidabilità delle metodologie d'analisi di laboratorio, i problemi di campionamento, oltre che gli aspetti teorici e le applicazioni della statistica e della genetica alla scienza forense, sono stati esaminati approfonditamente. In conclusione, oggi la genetica forense poggia su solide basi teoriche e metodologiche. Lo scopo fondamentale delle analisi di genetica forense è di verificare l'ipotesi che un DNA fingerprinting sia associato univocamente ad un particolare individuo, oppure che il DNA fingerprinting di un figlio sia derivato dai DNA fingerprinting dei due genitori presunti.

La Convenzione di Washington (CITES) è un trattato internazionale che ha lo scopo di disciplinare il commercio di piante ed animali. La distruzione degli ambienti naturali ed il commercio incontrollato di piante ed animali selvatici e dei loro prodotti, costituiscono le principali cause di rarefazione e di rischio d'estinzione di popolazioni e specie. La CITES si fonda sul principio che il controllo del commercio sostenibile degli animali e delle piante e dei loro prodotti, costituisca uno strumento di conservazione delle popolazioni selvatiche, soprattutto se il concetto dell'uso sostenibile delle specie viventi è posto alla base delle legislazioni internazionali e nazionali.

La corretta applicazione della CITES richiede, infatti, che la dinamica delle popolazioni di specie minacciate oggetto di traffico sia costantemente controllata. L'utilizzo sostenibile delle risorse naturali può anche

fornire profitti alle popolazioni umane locali, profitti che possono essere in parte utilizzati per programmi di conservazione.

L'attività della CITES si esplica tramite il rilascio di permessi d'importazione ed esportazione di individui vivi o dei loro prodotti, appartenenti a specie protette che sono iscritte alle Appendici I e II. Le specie di Appendice I godono di protezione totale ed il loro commercio e traffico internazionale sono strettamente limitati. Le specie di Appendice II possono sostenere certi livelli di prelievo e quindi il traffico è consentito in maniera controllata. La CITES disciplina anche la detenzione ed il traffico di piante ed animali riprodotti in cattività ed eventualmente utilizzati in esibizioni itineranti. In questi casi, i permessi CITES possono essere rilasciati solo se sia dimostrato che questi individui siano nati in cattività. Il regolamento CE 1808/2001 della Commissione delle Comunità Europee, relativo alla protezione di specie della flora e della fauna selvatiche mediante il controllo del loro commercio (Gazzetta Ufficiale CE n. L 250 del 19/09/2001), stabilisce che le autorità di gestione nazionali possano avvalersi di analisi genetiche per determinare l'origine e le relazioni di parentela di piante ed animali detenuti e riprodotti in cattività.

LA CONVENZIONE SUL COMMERCIO INTERNAZIONALE DELLE SPECIE ANIMALI E VEGETALI SELVATICHE MINACCIATE D'ESTINZIONE (CITES)

La CITES

La Convenzione sul commercio internazionale delle specie animali e vegetali selvatiche minacciate d'estinzione (Convention on International Trade in Endangered Species of Wild Fauna and Flora – CITES) è stata siglata a Washington il 3 marzo 1973 ed è entrata in vigore il 1° luglio 1975. La CITES è stata ratificata dallo Stato italiano con la legge del 19 dicembre 1975, n. 874 (Gazzetta Ufficiale 24 febbraio 1976, n. 49, S. O.) ed è stata depositata presso il Governo Svizzero, Depositario della Convenzione, il 2 ottobre 1979. La Convenzione, entrata in vigore per l'Italia il 31 dicembre 1979, è stata inizialmente siglata da 21 Stati. Attualmente più di 150 Stati sono membri della CITES. Inoltre l'Unione Europea, pur non essendo di per sé membro della CITES ha dettato norme uniformi per la sua applicazione fra gli Stati Membri dell'Unione attraverso alcuni Regolamenti Comunitari, dapprima con il Regolamento (CEE) n. 3626/82 del 3 dicembre 1982, entrato in vigore il 31 dicembre

1982 e con il suo regolamento di attuazione 3418/83 ed ora dal 1997 con il Regolamento (CE) 338/97 del Consiglio del 9 dicembre 1996 (Gazzetta Ufficiale CE n. L 61 del 03/03/1997) modificato da ultimo dal Regolamento (CE) 2704/2000 della Commissione del 30 Novembre 2000 (Gazzetta Ufficiale CE L320 del 18/12/2000).L'ultimo regolamento attuativo è il CE 1808/2001 della Commissione del 30 agosto 2001 (Gazzetta Ufficiale CE n. L 250, del 19/09/2001). Per un quadro completo dell'applicazione della CITES nell'Unione Europea è possibile consultare il sito della Commissione Europea (http://www.europa.eu.int/comm/environment/cites/home_en.htm)

Il Segretariato Generale CITES (<http://cites.org/>) è amministrato dall'UNEP (Programma delle Nazioni Unite per l'ambiente; <http://www.unep.org>) ed ha sede a Ginevra (Svizzera). Altri organismi internazionali che collaborano con il Segretariato CITES sono: gli uffici TRAFFIC (<http://www.traffic.org/>), organizzazione del WWF e della IUCN per l'analisi delle informazioni sul commercio internazionale di flora e fauna; la IUCN (<http://www.iucn.org>; Unione internazionale per la conservazione della natura); il WCMC (<http://www.unep-wcms.org/>; Centro mondiale per il monitoraggio della conservazione), che fornisce informazioni a sostegno delle politiche di conservazione della fauna e flora. Il WCNC ha i suoi uffici a Cambridge (UK) ed è parte integrale dell'UNEP.

Nel corso degli ultimi vent'anni la CITES ha rappresentato il più importante e, per certi aspetti, il principale strumento internazionale di controllo per la conservazione delle specie animali e vegetali minacciate o a rischio d'estinzione. Tramite gli strumenti legislativi offerti dalla CITES, infatti, in molti paesi si sono avviate azioni di controllo del commercio internazionale di piante ed animali o dei prodotti commerciali da loro derivati. Tramite la CITES si sono inoltre attivate azioni di monitoraggio dello status delle popolazioni di alcune specie minacciate. La CITES non è nata solo come strumento per tutelare la conservazione di specie-bandiera, vale a dire di specie particolarmente minacciate e di grande impatto sull'opinione pubblica. Queste specie, principalmente grandi erbivori e predatori, sono collocate ai vertici delle catene alimentari e svolgono ruoli cruciali nella regolazione del funzionamento di interi ecosistemi. La CITES intende anche tutelare una miriade di altre specie, apparentemente meno carismatiche, ma di straordinaria importanza per la conservazione dell'integrità e della funzionalità degli ecosistemi. L'estinzione di queste specie (ad es., chiroterri, coralli) avrebbe effetti negativi a cascata sull'intero ecosistema e potrebbe provocare gravi

crisi nella diversità biologica regionale. La CITES quindi assume un ruolo importante anche come strumento per la conservazione della biodiversità.

Di particolare rilevanza è il concetto di uso sostenibile delle risorse, introdotto dall'Articolo IV della CITES. L'Articolo IV dispone che il commercio internazionale di specie vegetali ed animali sia consentito solo in conformità ad un'attenta valutazione scientifica del ruolo che le specie oggetto di commercio hanno nell'ecosistema. Il prelievo non deve essere nocivo per la sopravvivenza della specie. In questo contesto è sempre più importante che i Paesi Consumatori, quelli del Nord del mondo, collaborino con i Paesi produttori di fauna (quelli del Sud) affinché le risorse siano utilizzate in modo razionale e sostenibile.

Il commercio di organismi viventi o dei loro prodotti è sostenuto da una serie di motivazioni, fra le quali assumono un peso particolarmente rilevante il commercio di piante ed animali vivi per gli allevamenti (sia a scopo commerciale sia amatoriale), il commercio di prodotti vegetali ed animali usati per il consumo o per la fabbricazione di oggetti; la compravendita di souvenir prodotti con parti di piante ed animali; l'uso di parti e principi attivi d'origine vegetale ed animale per la medicina tradizionale; i prodotti vegetali ed animali usati a scopo alimentare; gli scambi di campioni vegetali ed animali utilizzati per la ricerca scientifica; il commercio di selvaggina e di trofei di caccia. Tutte queste motivazioni possono apportare un beneficio alla conservazione delle specie se riportano ai Paesi di origine un vantaggio economico reinvestito nella conservazione degli ecosistemi.

Nel testo della CITES sono utilizzate, fra le altre, le seguenti definizioni (dall'Art. I della Legge 19 dicembre 1975, n. 874. Ratifica ed esecuzione della CITES. (Gazzetta Ufficiale 24 febbraio 1976, n. 49, S. O.):

- "Specie": ogni specie, sottospecie, oppure un gruppo di esseri viventi relativi alle medesime e geograficamente isolato;
- "Specimen ": qualsiasi animale o qualsiasi pianta, vivi o morti; ... ogni parte oppure ogni prodotto ottenuto dall'animale ... dalla pianta. Il termine è stato sostituito con "esemplare" nel regolamento CE n. 338/97 del 9 dicembre 1996;
- "Commercio": l'esportazione, la riesportazione, l'importazione e l'introduzione con provenienza dal mare;
- "Parte": uno Stato per il quale la CITES è entrata in vigore.

L'azione della CITES si basa sull'iscrizione alle Appendici I, II e III di specie che hanno diversi livelli di protezione (Art. I della CITES). In

particolare, è consentito il traffico delle specie in Appendice I, esclusivamente per scopi non commerciali, ad esempio scientifici. Le specie in Appendice II possono sostenere certi livelli di prelievo, e quindi il loro commercio è consentito in maniera controllata. La definizione del prelievo compatibile è effettuata anche tramite un sistema di quote, che richiede un monitoraggio centralizzato della dinamica delle popolazioni e del commercio dei singoli esemplari o dei loro prodotti.

Articolo II della CITES. Principi fondamentali.

- L'Appendice I elenca le specie (circa 675) maggiormente a rischio d'estinzione e ne proibisce il commercio, ad esclusione di particolari eccezioni, che devono essere comunque autorizzate tramite permessi d'importazione ed esportazione; il commercio delle specie d'Appendice I è di solito vietato, mentre gli scambi di campioni per scopi scientifici o gli scambi di individui fra giardini zoologici possono essere autorizzati;
- L'Appendice II elenca specie (circa 25 000) non ancora a rischio di estinzione, ma che sarebbero fortemente minacciate se il commercio non fosse controllato e comunque autorizzato dai permessi di importazione ed esportazione; inoltre elenca specie che pur non essendo direttamente minacciate appartengono a generi, famiglie o ordini che potrebbero essere confuse con specie di Appendice I (ad esempio, l'Appendice II include tutti i pappagalli);
- L'Appendice III elenca specie protette localmente, e consente alla comunità internazionale di collaborare al controllo del commercio di specie che godono di protezione a livello nazionale.

Una lista aggiornata delle specie vegetali ed animali iscritte alle Appendici CITES, è riportata nel Regolamento CE 2704/2000 della Commissione del 30 Novembre 2000 (Gazzetta Ufficiale CE L320 del 18/12/2000). Questo regolamento CE stabilisce che le specie elencate nelle Appendici CITES I, II e III siano incluse negli Allegati A, B, C e D, secondo il dettato dell'Articolo 3. Gli aggiornamenti alle Appendici CITES sono costantemente riportati nel sito CITES (<http://cites.org/>). Gli aggiornamenti ai Regolamenti Comunitari sono riportati nel sito della Commissione Europea, citato in precedenza.

Le linee-guida ed i criteri per l'iscrizione (o la cancellazione) delle specie alle Appendici furono definiti, per la prima volta, nel corso della Conferenza di Berna (1976). Al fine di tener conto dei progressi in biologia della conservazione, i criteri di Berna furono aggiornati dalla Nona Conferenza delle Parti, con l'approvazione della Risoluzione Conf.9.24,

che ebbe l'apporto tecnico dell'Animals Committee e del Plants Committee, le cui riunioni si svolgono periodicamente. Questi criteri sono ora in fase di ulteriore revisione, che sarà discussa nella prossima riunione della Conferenza delle Parti (Santiago del Cile, novembre 2002).

La Conferenza delle Parti (CoP) redige ed approva delle risoluzioni, che indicano linee-guida e raccomandano specifiche azioni agli Stati membri della CITES. L'inclusione delle specie nelle tre Appendici vincola le parti ad applicare specifici controlli sull'importazione e l'esportazione. I singoli Stati emettono regolamenti in applicazione della CITES. Le legislazioni nazionali devono essere basate anche sulle risoluzioni approvate dalla CoP.

Gli Stati partner devono istituire una o più Autorità di gestione ed un'Autorità scientifica, le quali operano autonomamente l'una dall'altra. L'Autorità di gestione rilascia i necessari permessi CITES e compila le informative annuali per la CITES, partecipa ai lavori delle CoP, ecc. L'Autorità scientifica, tramite le delibere della Commissione Scientifica CITES (CSC) determina se il commercio di una particolare specie sia dannoso alla sua sopravvivenza, controlla i volumi di commercio nel rispetto delle quote e ne valuta gli impatti sulle popolazioni naturali, verifica che le condizioni di detenzione in cattività di specie CITES siano adeguate alle esigenze di benessere degli animali.

L'attività delle Autorità CITES nazionali si avvale della collaborazione delle ONG (ad es., gli uffici TRAFFIC del WWF) e delle istituzioni scientifiche (Università e Musei; associazioni degli Zoo ed Acquari, ecc.). In Italia si sono istituite tre Autorità di gestione: la principale è il Ministero dell'Ambiente e della Tutela del Territorio; altri organi di gestione sono il Ministero per le Politiche Agricole, Divisione II, il Servizio CITES del Corpo Forestale dello Stato ed il Ministero delle Attività Produttive. L'Autorità scientifica ha sede presso la Divisione II della Direzione Conservazione della natura del Ministero dell'Ambiente e della Tutela del Territorio.

L'Art. VII della CITES elenca le possibili eccezioni alle regolamentazioni dei permessi così come sono definite dalla Convenzione. Le eccezioni consentite dalla CITES riguardano il commercio di: esemplari pre-Convenzione; effetti personali; animali nati in cattività; piante propagate artificialmente; scambi di esemplari fra istituzioni scientifiche; piante ed animali presenti in zoo, circhi ed altre esibizioni itineranti. Tutte queste eccezioni devono essere chiaramente definite e specificate dalle legislazioni nazionali al fine di impedire illegalità. Ogni altra eccezione costituirebbe una violazione della CITES. Le parti possono sempre

applicare misure di controllo più stringenti di quelle richieste dalla CITES (Art. XIV.1). Le definizioni attualmente in vigore, sono quelle adottate dal regolamento CE n. 338/97 del 9 dicembre 1996 (Articolo 2) e dal recente regolamento CE 1808/2001.

L'Art. VII.4 della CITES disciplina il commercio di animali riprodotti in cattività. Gli esemplari di specie animali d'Appendice I riprodotti in cattività per scopi commerciali, devono essere considerati e trattati come esemplari di specie incluse nell'Appendice II. I permessi richiesti per questi esemplari sono equivalenti ai permessi applicabili agli esemplari d'Appendice II. In particolare, non sono richiesti permessi d'importazione dallo Stato d'importazione. Gli esemplari d'Appendice I riprodotti in cattività per scopi non commerciali, e d'Appendici II e III, riprodotti in cattività per qualsiasi scopo, possono essere scambiati liberamente, senza che sia richiesto alcun permesso, purchè sia dimostrato che questi esemplari sono nati in cattività (Art. VII.5). La Risoluzione Conf 10.16 (Rev) riguarda il trattamento degli animali nati in cattività. A livello comunitario queste procedure sono state riprese nei Regolamenti Comunitari 338/97 e 1808/1. Quest'ultimo regolamento sostituisce il Regolamento 939/97 ora abrogato.

La riserva riproduttiva deve essere costituita senza nuocere alla sopravvivenza della specie in natura. Allo scopo di prevenire l'inbreeding negli stock riprodotti in cattività, è possibile ottenere l'autorizzazione ad effettuare prelievi limitati dalle popolazioni naturali. La riserva riproduttiva deve essere gestita in modo da garantirne il mantenimento perenne.

In conseguenza di queste norme, l'Autorità di gestione può rilasciare i permessi d'esportazione a scopi commerciali di esemplari d'Appendice I riprodotti in cattività, solo dopo avere accertato che gli esemplari siano effettivamente nati in cattività e che le condizioni dettate dalla risoluzione Conf. 10.16 (Rev) e riprese dall'articolo 24 del regolamento 1808/2001, siano rispettate. Le procedure di standardizzazione per il rilascio dei certificati CITES di riproduzione in cattività, sono state definite dalla risoluzione Conf 10.2 e dall'ultimo regolamento CE 1808/2001. Queste risoluzioni raccomandano che le parti indichino nei certificati di nascita in cattività l'origine degli esemplari, vale a dire se si tratti di esemplari d'Appendice I riprodotti in cattività per scopi commerciali, per scopi non commerciali, oppure se si tratti di esemplari d'Appendici II o III, o di esemplari di prima generazione (F1), nati in cattività, ma che non corrispondono ai termini del citato articolo 24. Le stesse raccomandazioni si applicano alle parti ed ai prodotti di questi esemplari. I certificati di nascita in cattività devono sempre includere il nome scientifico della

specie cui appartiene l'esemplare, il numero della marcatura ed il numero di registro dell'operazione commerciale. Per gli animali nati in cattività, in conclusione, le legislazioni nazionali devono dichiarare esplicitamente che gli esemplari d'Appendice I nati in cattività richiedono permessi d'esportazione per scopi commerciali; richiedere che la stessa certificazione sia rilasciata anche per tutti gli altri esemplari d'Appendice I che siano nati in cattività; esplicitare le procedure per il rilascio dei certificati richiesti dalle operazioni commerciali per specie d'Appendice I. Inoltre, queste procedure devono includere i criteri per l'identificazione individuale degli esemplari e le procedure per applicare tutte le altre forme di controllo delle nascite in cattività.

Esiste evidentemente il rischio che esemplari prelevati dalle popolazioni naturali siano immessi nei circuiti commerciali come se fossero nati in cattività. L'Autorità di gestione, prima di rilasciare i certificati, è tenuta ad ottenere prove conclusive che gli esemplari siano prodotti di seconda generazione in cattività. Queste prove possono anche essere fornite dai risultati di analisi genetiche.

La CoP ha deciso (risoluzione Conf. 4.15) di stabilire un registro di tutte le operazioni commerciali di specie d'Appendice I riprodotte in cattività. Il registro deve naturalmente imporre condizioni di controllo sulle licenze, quindi gli esemplari devono essere marcati, le operazioni commerciali possono essere ispezionate e le licenze possono essere revocate. L'Articolo XIV.1 autorizza l'applicazione dei controlli richiesti dall'Appendice I ad ogni esemplare di specie d'Appendice I nato in cattività, e non solo agli esemplari che sono riprodotti per scopi commerciali.

L'Art. VII.4 della CITES afferma che gli esemplari di specie vegetali d'Appendice I, propagati artificialmente per scopi commerciali, devono essere considerati e trattati come esemplari di specie incluse nell'Appendice II. La risoluzione Conf. 11.11 raccomanda che il termine "propagate artificialmente" sia riferito solo alle piante cresciute dall'uomo utilizzando semi, tessuti del callo, spore od altri propaguli in condizioni controllate. La riserva riproduttiva delle piante propagate artificialmente deve essere costituita e mantenuta senza nuocere alla sopravvivenza della specie in natura. La riserva riproduttiva deve essere gestita in modo da garantirne il mantenimento perenne.

I controlli da esercitare sulle operazioni commerciali che propagano artificialmente piante, ed i certificati d'esportazione di piante propagate artificialmente devono essere regolati da procedure analoghe a quelle stabilite per gli animali nati in cattività. I certificati di propagazione artificiale devono includere le stesse informazioni richieste per gli animali nati

in cattività, e devono indicare l'origine degli esemplari di cui si richiede la certificazione. I permessi richiesti per questi esemplari sono equivalenti ai permessi applicabili agli esemplari d'Appendice II. In particolare, non sono richiesti permessi d'importazione dallo Stato d'importazione. Gli esemplari d'Appendice I che sono stati riprodotti in cattività per scopi non commerciali, e d'Appendici II e III riprodotti in cattività per qualsiasi scopo, possono essere scambiati liberamente, senza che sia richiesto alcun permesso, purché sia dimostrato che questi esemplari siano nati in cattività (Art. VII.5). I certificati devono quindi indicare esplicitamente se gli esemplari appartengono a piante d'Appendice I, oppure a loro parti o prodotti, che siano state propagate artificialmente per scopi commerciali o non commerciali, oppure se appartengono a specie d'Appendici II o III.

L'Art. VII.7 stabilisce che gli zoo, i circhi ed altre esibizioni itineranti di animali e piante possono movimentare senza permessi o certificati gli esemplari in loro possesso, purché gli esemplari siano detenuti legalmente, in quanto esemplari pre-Convenzione, oppure nati in cattività o propagati artificialmente. Le esportazioni e importazioni di questi esemplari devono essere controllate dalle Autorità di gestione competenti, che devono anche accertarsi che gli animali vivi siano trasportati ed ospitati nel rispetto delle loro esigenze di benessere, minimizzando il rischio di danni e di malattie.

L'Art. VI.7 definisce le procedure di marcaggio. Ogni qualvolta sia appropriato e possibile l'Autorità di gestione dovrebbe provvedere affinché gli esemplari siano marcati con sistemi che ne consentano l'identificazione. La legislazione dovrebbe quindi dotare l'autorità competente dei regolamenti che consentano il marcaggio degli esemplari CITES. Queste procedure sono d'importanza particolare per gli esemplari pre-Convenzione, per gli animali nati od ospitati in cattività, per gli esemplari importati legalmente o prelevati dalla natura, per gli esemplari di specie sottoposte a quote d'esportazione, e per esemplari in esibizioni itineranti. La CoP ha adottato numerose Risoluzioni che indicano quali categorie di esemplari dovrebbero essere marcati e come: gli animali in recinti (Conf. 5.16) e gli animali nati in cattività. In particolare, si prescrive l'inanellamento degli uccelli di specie in Appendice I (Conf. 6.21); il marcaggio di animali vivi con microchips (Conf. 8.13); il marcaggio di animali vivi esibiti nelle mostre (Conf. 8.16). Il regolamento CE 1808/2001 stabilisce ulteriori dettagli per eseguire le procedure di marcaggio.

Sequestro e confisca degli esemplari detenuti illegalmente. Il sequestro è una misura temporanea che può essere presa dalle autorità nazionali incaricate di applicare le disposizioni della CITES, nell'attesa delle deci-

sioni definitive nel merito dei casi specifici. La decisione definitiva può comportare la confisca degli esemplari o la restituzione ai loro legittimi proprietari. La decisione di confisca non richiede, normalmente, la prova che l'esemplare sia commerciato o posseduto illegalmente. Devono sussistere, tuttavia, ragionevoli sospetti. Ogni qualvolta l'autorità competente abbia il sospetto che un esemplare sia stato importato, esportato, detenuto o commerciato illegalmente, può disporre il sequestro. E' anche possibile disporre il sequestro dei discendenti o dei propaguli di ogni animale o pianta sequestrata. E' possibile che gli esemplari sequestrati siano affidati provvisoriamente al proprietario. In questo caso i costi per la custodia ed il mantenimento degli esemplari sequestrati devono essere sostenuti dai proprietari. Gli esemplari sequestrati che sono stati abbandonati dai loro proprietari o da ignoti, le piante o gli animali che sono deceduti mentre erano tenuti in custodia a seguito di sequestro, devono essere collocati a discrezione dell'autorità competente. Tutti i costi (custodia, trasporto, collocazione degli esemplari non viventi, mantenimento degli esemplari viventi) sostenuti nel corso di un sequestro devono essere considerati come un debito dello Stato, debito che può essere recuperato a carico dell'importatore o dello spedizioniere degli esemplari sequestrati. La Risoluzione Conf.10.7 stabilisce le linee guida per la gestione degli animali confiscati. Tali linee-guida dovrebbero essere adottate nelle legislazioni nazionali degli Stati Membri.

L'Art. VIII.1 impone la confisca o la restituzione allo Stato d'origine degli esemplari commerciatati illegalmente. La confisca può essere imposta tramite una sentenza emessa da una corte di giustizia, o da un ordine dell'autorità amministrativa. In Italia la confisca è di natura penale e amministrativa, così come stabilito dalla Legge 7 febbraio 1992 n. 150. Gli esemplari confiscati diventano proprietà dello Stato che può disporre come ritiene giusto. Gli oggetti possono essere venduti tramite asta pubblica. Tuttavia, la vendita di esemplari d'Appendice I violerebbe lo spirito della Convenzione, e quindi in questo caso sono necessarie regole particolari (Art. VIII.4). Un esemplare vivo confiscato dovrebbe essere affidato all'Autorità di gestione dello Stato che ha operato la confisca. L'Autorità di gestione dovrebbe provvedere, dopo adeguate consultazioni con lo Stato d'origine, a restituire l'esemplare a quello Stato a spese dello Stato, oppure dovrebbe affidare l'esemplare ad un centro d'accoglienza e di recupero, o ad altra struttura analoga. Queste strutture possono usare gli esemplari confiscati esclusivamente per quegli scopi non commerciali, scientifici o educazionali che possano contribuire alla sopravvivenza della specie. L'Art. VIII.5 definisce un centro d'accoglienza e recupero come

un'istituzione designata dall'Autorità di gestione a prendersi cura del benessere degli esemplari vivi che sono stati confiscati. Gli esemplari d'Appendice I restituiti agli Stati d'origine possono essere rilasciati in natura, oppure trasferiti ad un centro di recupero o ad altra struttura analoga, che li potrà usare esclusivamente per quegli scopi non commerciali, scientifici o educazionali, che possano contribuire alla sopravvivenza della specie. La risoluzione Conf. 4.18 raccomanda che i costi di confisca, custodia e restituzione degli esemplari d'Appendice II agli Stati d'origine, siano a carico di chi ha violato la legge.

Gli esemplari confiscati e morti di specie d'Appendice I dovrebbero essere affidati ad istituzioni scientifiche riconosciute ed utilizzati esclusivamente per scopi scientifici, educazionali o per identificazione, oppure altrimenti distrutti (risoluzione Conf. 3.14). Naturalmente, le persone che hanno violato la legge non devono ottenere alcun vantaggio finanziario o altro dalla confisca ed affidamento di esemplari vivi o morti.

La ri-esportazione verso il paese d'origine degli esemplari d'Appendice I confiscati può essere concessa, assumendo che tali esemplari siano stati importati legalmente (risoluzione Conf. 9.10 - Rev). Tuttavia la CoP suggerisce che gli esemplari confiscati debbano essere restituiti se e solo se gli Stati d'origine ne abbiano specificatamente fatto richiesta e siano preparati ad affrontare i costi della restituzione. La stessa risoluzione Conf. 9.10 (Rev.) raccomanda anche che gli esemplari vivi d'Appendice I siano restituiti agli Stati d'origine, quando ciò sia pratico e di beneficio per la conservazione della specie. La stessa risoluzione raccomanda che gli esemplari vegetali vivi siano restituiti agli Stati d'origine se possono essere reinseriti negli ambienti naturali, o se possono essere usati per la propagazione artificiale. La risoluzione inoltre raccomanda che gli animali d'Appendici II e III confiscati vivi siano restituiti, quando possibile ed appropriato, alle Autorità di controllo degli Stati d'esportazione o ri-esportazione o d'origine. Questa risoluzione chiarisce che possono essere restituiti agli Stati d'origine anche esemplari vivi che non siano stati ancora confiscati. Infatti, l'Art. VIII.b della CITES consente alle Parti di scegliere se confiscare o restituire immediatamente gli esemplari vivi importati illegalmente.

In Italia sono in fase di costituzione alcuni centri di accoglienza pubblici e privati, destinati ad accogliere esemplari di specie CITES che sono stati confiscati. L'identificazione di specie, di individui e le diagnosi di parentela vengono effettuate anche tramite il supporto di analisi genetiche. L'esecuzione delle analisi di genetica forense in applicazione della CITES in Italia vengono disposte dall'Autorità di gestione che ha sede

presso il Ministero per le Politiche Agricole, Divisione II, Servizio CITES del Corpo Forestale dello Stato. Le analisi di genetica forense attualmente vengono eseguite in prevalenza presso il Laboratorio di genetica dell'Istituto Nazionale per la Fauna Selvatica.

STRUTTURA E FUNZIONI DEL DNA

Genetica forense e DNA fingerprinting

La genetica molecolare ha sviluppato metodi d'analisi del DNA che consentono l'identificazione di ogni individuo presente in una popolazione e la ricostruzione delle relazioni di parentela all'interno dei nuclei familiari. I risultati delle analisi del DNA producono informazioni che sono utilizzabili come prova di accusa o difesa nei procedimenti legali e nelle aule dei tribunali. Le procedure di genetica forense devono garantire risultati di qualità elevata, che devono essere valutati accuratamente e devono risultare comprensibili anche a chi non è genetista di professione. Le analisi di genetica forense vengono realizzate per fornire alle autorità competenti informazioni obiettive che possono aiutare a prendere decisioni o a risolvere dispute legali. Tuttavia, la scienza forense non stabilisce chi è innocente o colpevole, chi ha violato la legge oppure no. La scienza forense fornisce informazioni che servono a ricostruire gli eventi e le azioni. Non giudica se determinate azioni siano legali o no. Le ricostruzioni della scienza forense avvengono essenzialmente tramite associazioni: un particolare tipo di DNA, che possiamo definire come genotipo, o "DNA fingerprinting", ricavato da uno o più campioni biologici, viene associato ad un determinato individuo. In questo senso le analisi del DNA consentono di "identificare" i campioni biologici. I metodi di analisi molecolare che consentono di ricostruire i "DNA fingerprinting" si basano sulla osservazione di sistemi di frammenti di DNA molto complessi e molto variabili, che sono associati univocamente ad ogni individuo. Un oggetto (nel nostro caso un campione biologico) è identificato quando è collocato in una classe di oggetti che hanno caratteristiche simili. Ovviamente, ogni classificazione include nella stessa classe alcuni oggetti che sono fra di loro simili e, contemporaneamente, esclude altri oggetti che sono fra di loro meno simili e che dovranno essere collocati in altre classi. Vedremo in seguito, che la genetica forense esegue procedure di "identificazione" dei campioni biologici sulla base di una logica strettamente probabilistica. Quando le caratteristiche di un oggetto sono

uniche, allora l'oggetto è individualizzato e la sua classe d'appartenenza escluderà tutti gli altri oggetti. Le analisi del DNA consentono di "individualizzare" i campioni biologici, proprio perché ogni individuo è geneticamente unico, ad eccezione dei gemelli monozigoti.

Il valore dei metodi di genetica forense dipende dalla possibilità di generare evidenze che consentano l'identificazione e l'individualizzazione dei campioni biologici, oltre che dalla valutazione del grado di associazione che esiste fra diversi campioni. Le impronte digitali sono percepite dall'opinione pubblica e dalla legge come tratti convincenti di individualità. Il loro valore di tratti individuali è stato sempre valutato empiricamente, poiché l'esame delle impronte digitali in decine di migliaia di persone non ha mai portato alla scoperta di impronte identiche appartenenti a persone diverse. La struttura delle impronte digitali dipende dall'azione di molteplici fattori genetici e non genetici che si esprimono durante lo sviluppo embrionale di ogni persona. Perciò, anche due gemelli monozigoti hanno impronte digitali differenti. L'affermazione che: "due impronte digitali umane sono identiche, quindi sono state lasciate dallo stesso dito di una mano, quindi appartengono alla stessa persona" è accettata senza discussione come "prova" dal valore indiscutibile, anche se non esistono forti giustificazioni biologiche o statistiche che possano sostenerla. Al contrario, la struttura del DNA fingerprinting è determinata da mutazioni genetiche di geni che sono, quasi sempre, ben identificati. La variabilità del DNA fingerprinting è analizzata rigorosamente utilizzando i modelli della genetica delle popolazioni e le procedure della statistica. L'utilizzo della genetica molecolare nella scienza forense si basa su forti giustificazioni biologiche e statistiche.

Prima dello sviluppo della genetica molecolare sono stati utilizzati in genetica forense altri metodi di analisi della variabilità biologica, come ad esempio la determinazione dei gruppi sanguigni, dei polimorfismi proteici ed, in particolare, degli alloenzimi. Questi sistemi genetici sono analizzati utilizzando campioni di sangue. Dopo lo sviluppo delle metodiche di analisi della genetica molecolare, questi metodi di analisi sono stati progressivamente abbandonati. La superiorità delle analisi del DNA è molteplice: il DNA è una molecola molto più stabile di qualsiasi proteina o enzima; esistono tecniche che consentono di amplificare le più minute tracce di DNA; la variabilità genetica presente nelle sequenze di DNA è enorme. Perciò ogni individuo possiede un patrimonio genetico unico che può essere descritto usando i più appropriati fra i molti metodi di analisi disponibili.

Introduzione al DNA fingerprinting

Ogni individuo, eccetto i gemelli monozigoti, è geneticamente unico, nel senso che possiede un patrimonio di informazioni genetiche unico. Queste informazioni sono scritte nel DNA del genoma individuale e possono essere visualizzate utilizzando le tecniche di analisi della genetica molecolare. Il concetto di DNA fingerprinting deriva dai metodi, ampiamente usati in criminologia, di identificazione delle impronte digitali. Il DNA fingerprinting è l'impronta genetica di ciascun individuo. I DNA fingerprinting sono ampiamente utilizzati in genetica forense ed in criminologia, e sono applicati alla risoluzione di test di paternità, identificazione di specie ed individui di piante ed animali, bracconaggio e traffico di esemplari viventi e dei loro prodotti. I test di DNA fingerprinting possono ridurre considerevolmente i margini di soggettività che sono insiti in tutte le procedure di identificazione, purché essi siano eseguiti e valutati correttamente.

Il patrimonio genetico (genoma) di ogni individuo è unico. Tutte le cellule che costituiscono il corpo di un individuo contengono lo stesso identico genoma, quindi il DNA può essere estratto da qualsiasi tessuto (campioni di sangue e di tessuti solidi, biopsie, radici di capelli, peli, penne, frammenti ossei, saliva, escrementi, unghie, ecc.). L'unicità individuale e l'identità delle sequenze del DNA in qualsiasi tessuto del corpo di ciascun individuo forniscono le basi per il DNA fingerprinting. I DNA fingerprinting si ottengono applicando svariati metodi di analisi che sono stati definiti come test d'identità, profili genetici, tipizzazione genica, genotipizzazione, ecc. I concetti di "DNA fingerprinting" e di "profilo genetico individuale" sono sostanzialmente equivalenti.

L'idea e la storia del DNA fingerprinting nascono nel 1985, a seguito dei lavori di A. Jeffreys e collaboratori, i quali descrissero metodi di identificazione ed analisi di sequenze di DNA ripetute ed ipervariabili presenti nel genoma umano (Jeffreys *et al.* 1985). Jeffreys e collaboratori identificarono una sequenza di DNA, lunga 33 nucleotidi, ripetuta quattro volte all'interno del gene della mioglobina umana. Ognuna di queste sequenze ripetute ("repeat") contiene un modulo di 16 nucleotidi, costituito da una sequenza conservata (sequenza "core") che è stata, in seguito, identificata anche in molte altre sequenze di DNA ripetitivo. Queste sequenze di DNA ripetitivo, chiamate "minisatelliti", sono presenti in numerose copie nei cromosomi del genoma umano e di quasi tutte le altre specie viventi. I minisatelliti sono ipervariabili, perché il numero di ripetizioni dei repeat cambia frequentemente. Il genoma di ogni individuo possiede una combinazione unica di sequenze ripetute. L'identi-

ficazione dei minisatelliti consente perciò di ricostruire profili genetici individuali, i DNA fingerprinting. Se i DNA fingerprinting ottenuti da due differenti campioni biologici sono identici, è molto probabile che essi appartengano allo stesso individuo. Le sequenze ripetute sono trasmesse dai genitori ai figli secondo gli schemi della semplice eredità mendeliana. Perciò, se le sequenze ripetute di un figlio sono presenti anche nei due presunti genitori, è molto probabile che essi siano i suoi genitori naturali. Tuttavia, una corretta interpretazione dei DNA fingerprinting richiede una precisa conoscenza delle leggi dell'ereditarietà e della genetica delle popolazioni. I risultati delle analisi di DNA fingerprinting devono essere valutati ed interpretati ricorrendo ad appropriati strumenti di teoria delle probabilità e di analisi statistica.

Il concetto di "DNA fingerprinting" è usato per descrivere svariate tecniche di analisi genetica, inclusi i metodi che si basano sulla PCR per l'amplificazione casuale di frammenti polimorfici di DNA (Random Amplified Polymorphic DNA – RAPD; Amplified Fragment Length Polymorphism - AFLP). L'estensione del termine DNA fingerprinting a queste tecniche è, tuttavia, ingiustificato. La caratteristica fondamentale del DNA fingerprinting è quella di rivelare combinazioni di frammenti di DNA (alleli) uniche e distinte per ogni individuo, e quindi di consentire l'individualizzazione di ogni campione. Normalmente, due individui presi a caso da una popolazione condividono meno del 50% dei frammenti presenti nei loro rispettivi DNA fingerprinting. Questi frammenti sono ereditati in modo mendeliano, sono codominanti, metà dei frammenti sono ereditati dalla madre e metà dal padre. Questo non è sempre vero per le altre tecniche, che, pur rivelando spesso ampia variabilità, sia entro popolazione che tra popolazioni, non sono però sempre in grado di distinguere un individuo dall'altro. Inoltre, metodi come il RAPD e l'AFLP evidenziano frammenti di DNA fra cui esistono relazioni di dominanza, il che rende problematica la descrizione della variabilità nei profili genetici individuali. E' quindi opportuno limitare la definizione di "DNA fingerprinting" a quei metodi di analisi molecolare che consentono l'individualizzazione dei campioni. Questi metodi includono: la tipizzazione dei DNA fingerprinting multilocus classici, realizzati tramite sonde di DNA multilocus (multi-locus probes – MLP); la tipizzazione di loci multipli tramite DNA fingerprinting a locus singolo (questi loci sono detti variable number tandem repeats – VNTR), tramite l'uso di sonde di DNA specifiche per singoli loci (single locus probes – SLP); l'analisi di loci microsatelliti tramite PCR (questi loci sono anche chiamati short tandem repeats – STR). Independentemente dal metodo

utilizzato, il sistema dei frammenti di DNA che sono identificati nel campione costituisce il “profilo genetico” dell’individuo da cui proviene il campione.

Struttura e funzioni del DNA

Gli organismi eucarioti con genoma diploide sono costituiti da cellule il cui nucleo e citoplasma sono separati da una membrana cellulare, e sono dotati di coppie di cromosomi, metà dei quali sono ereditati dal padre e metà dalla madre (Fig. 1). Il corredo cromosomico di una cellula è detto “cariotipo diploide” ($2n$). Il DNA è organizzato nei cromosomi che sono contenuti nel nucleo cellulare (DNA nucleare), e nei mitocondri, organelli presenti nel citoplasma cellulare (DNA mitocondriale, mtDNA) (Fig. 2). La maggior parte delle cellule contiene il nucleo e i mitocondri, ad eccezione dei globuli rossi dei mammiferi, che non contengono il nucleo. Il DNA è organizzato in forma di doppia elica, costituita da due catene lineari di quattro nucleotidi (adenina - A; timina - T; guanina - G e citosina - C; Fig. 3). La struttura a doppia elica fu descritta da Watson e Crick nel 1953. Le catene di DNA sono complementari, perché possono appaiarsi seguendo una precisa regola: A=T (adenina e timina si legano costituendo due legami idrogeno) e C=G (citosina e guanina si legano costituendo tre legami idrogeno). L’ordine lineare con cui i quattro nucleotidi si susseguono nella doppia elica del DNA è detta sequenza

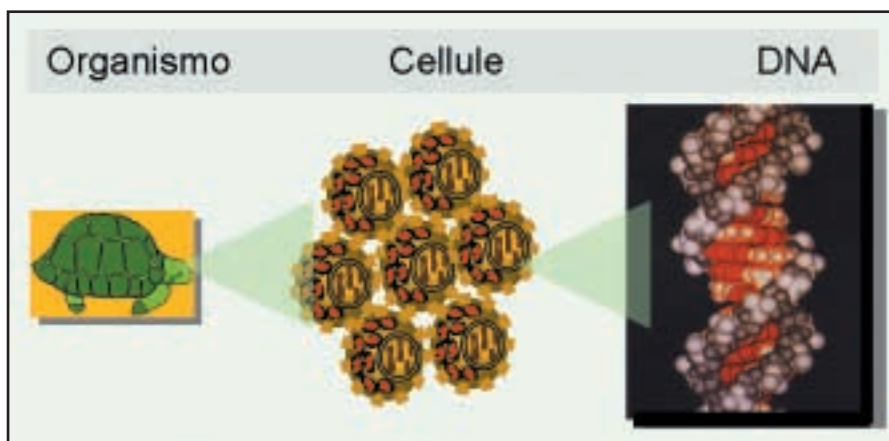


Figura 1 - Gli organismi eucarioti con genoma diploide sono costituiti di cellule dotate di nucleo e citoplasma. Il DNA è contenuto nel nucleo cellulare e nei mitocondri.

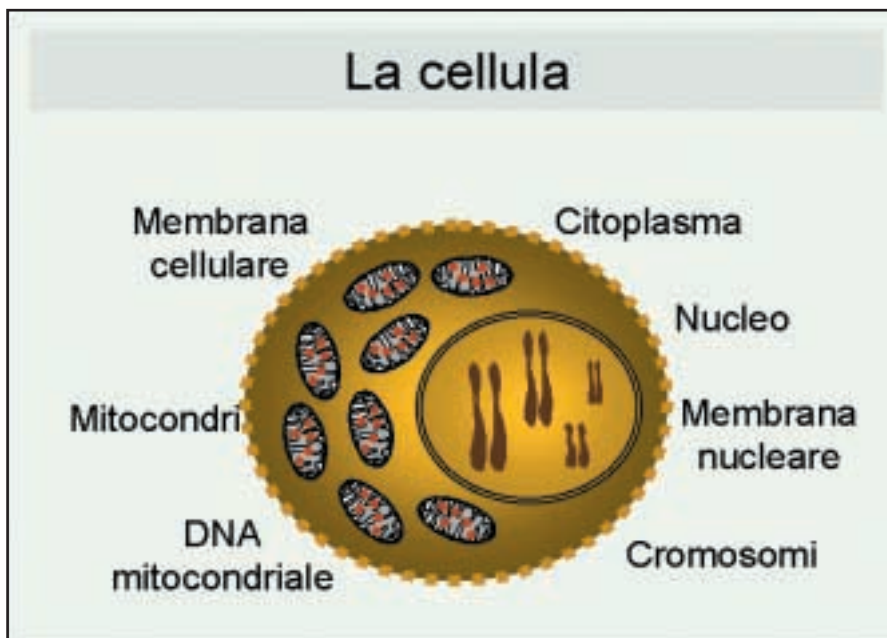


Figura 2 - Ogni cellula è delimitata dalla membrana cellulare e contiene il citoplasma e svariati organelli cellulari. Il materiale genetico (DNA) è contenuto nel nucleo cellulare (DNA nucleare) e nei mitocondri (DNA mitocondriale, mtDNA). Nucleo e mitocondri sono delimitati dalle rispettive membrane.

nucleotidica (Fig. 4). La doppia elica è una struttura molto semplice, estremamente stabile, e consente al DNA di svolgere le funzioni di substrato informativo per la sintesi proteica e di autoreplicazione. Il meccanismo della sintesi proteica è alla base dell'espressione funzionale e fenotipica dell'informazione genetica. Il meccanismo della replicazione è alla base della trasmissione ereditaria dell'informazione genetica.

Il DNA viene replicato prima che ogni divisione cellulare sia completata. Ciascuna delle cellule figlie riceve un nuovo set completo di cromosomi. Ognuna delle due catene di DNA (cromatidi) è replicata quando il DNA viene denaturato e la doppia elica si apre in un punto (Fig. 5). L'enzima che catalizza la replicazione, la DNA polimerasi, si aggancia nella zona di denaturazione ed avvia la ricopiatura controllando l'inserzione dei nucleotidi. Ogni cromatidio parentale funge da stampo per la sintesi di un nuovo cromatidio figlio, che si forma seguendo le regole di appaiamento molecolare di Watson e Crick. Le due nuove doppie eliche sono identiche, ciascuna essendo formata da un cromatidio

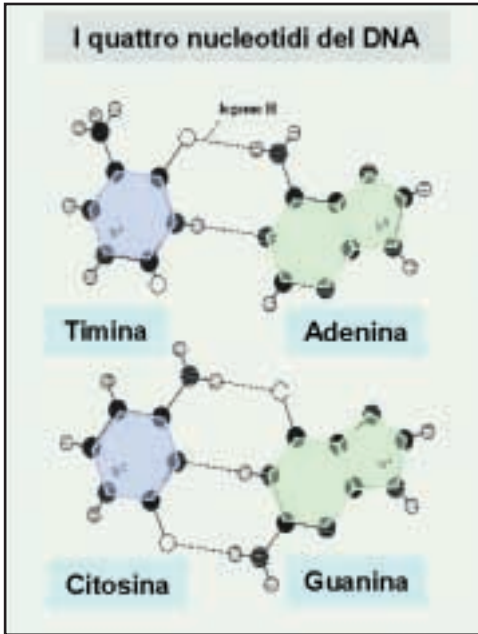


Figura 3 - Il DNA è composto da due catene lineari di quattro nucleotidi: adenina (A), timina (T), guanina (G) e citosina (C), ed è organizzato in forma di doppia elica. La doppia elica fu descritta da Watson e Crick nel 1953. Le catene di DNA sono complementari, perché possono appaiarsi seguendo una precisa regola: $A=T$, che si legano costituendo due legami idrogeno, e $C=G$, che si legano costituendo tre legami idrogeno (regole di appaiamento molecolare di Watson e Crick).

parentale e da un cromatidio complementare (Fig. 6). In questo modo, le sequenze di DNA sono copiate fedelmente e le informazioni genetiche codificate nelle sequenze sono conservate nel corso delle duplicazioni cellulari. Il meccanismo di copiatura non è perfetto ed alcune mutazioni nucleotidiche possono essere inserite casualmente. Le mutazioni modificano le sequenze di DNA e generano variabilità genetica. Le cellule, e quindi il DNA, si dividono continuamente nel corso dello sviluppo e della vita di un organismo. La divisione delle cellule dei tessuti somatici è chiamata mitosi (Fig. 7) e non ha alcuna implicazione per la trasmissione ereditaria dell'informazione genetica alla generazione successiva. Ogni coppia di cellule originatasi da ogni divisione cellulare somatica, contiene esattamente lo stesso DNA

della cellula parentale. Perciò, la determinazione del DNA fingerprinting di un individuo può essere effettuata utilizzando campioni di DNA estratto da qualsiasi tessuto e fornisce identici risultati.

Nel corso della formazione dei gameti (Fig. 8), nelle cellule della riproduzione (spermatozoi e cellule-uovo), il contenuto del DNA delle cellule somatiche diploidi viene dimezzato e la cellula diventa aploide (n). Questo meccanismo di divisione cellulare è chiamato meiosi (Fig. 7). La riduzione meiotica del corredo cromosomico è essenziale per conservare inalterato nel corso della fecondazione il numero diploide di cromosomi che è tipico di ogni specie. Quando la cellula-uovo viene fecondata da



Figura 4 - Il DNA è organizzato nei cromosomi, che sono contenuti nel nucleo cellulare. Ogni gene occupa una posizione precisa (locus genico) nei cromosomi. Il DNA è organizzato in forma di doppia elica. L'ordine lineare con cui i quattro nucleotidi si susseguono nella doppia elica del DNA è detta sequenza nucleotidica. Ogni gene codifica per una o più proteine.

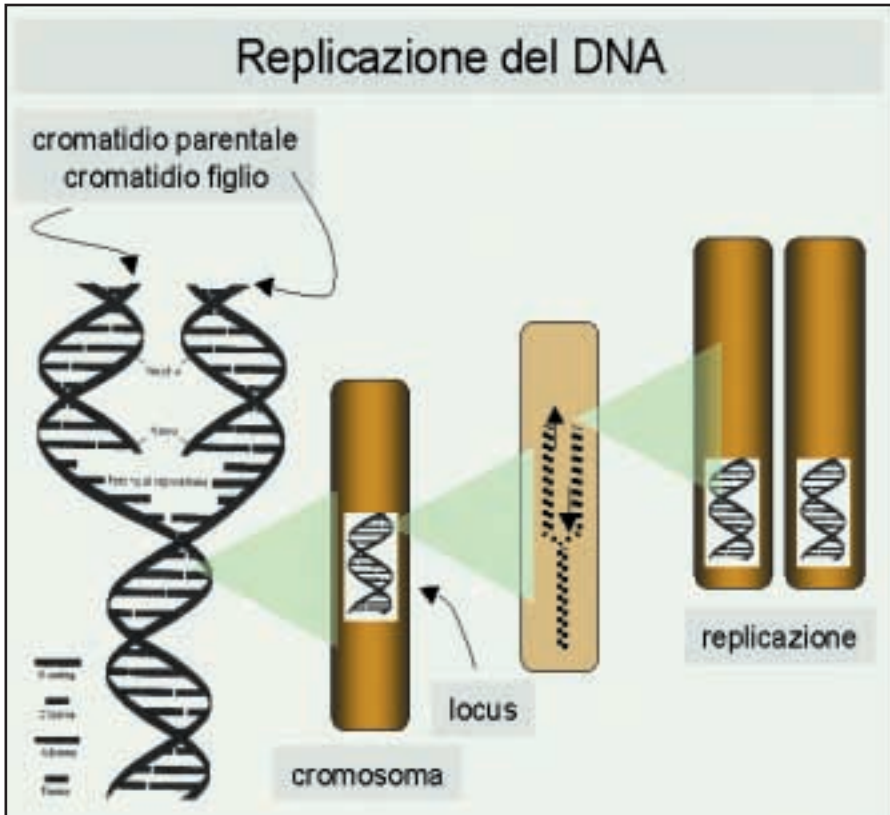


Figura 5 - Il DNA è replicato prima che ogni divisione cellulare sia completata. Ognuna delle due catene di DNA (cromatidi) è replicata quando il DNA viene denaturato e la doppia elica si apre in un punto predeterminato. Ciascuna delle due cellule figlie riceve un nuovo set completo di cromosomi.

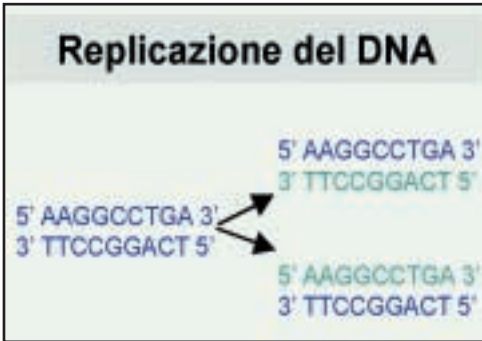


Figura 6 - Nel corso della replicazione del DNA ogni cromatidio parentale funge da stampo per la sintesi di un nuovo cromatidio figlio, che si forma seguendo le regole di appaiamento molecolare di Watson e Crick. Le due nuove doppie eliche sono identiche, ciascuna essendo formata da un cromatidio parentale e da un cromatidio complementare.

uno spermatozoo, il nucleo cellulare materno e quello paterno si uniscono ed i due complementi cromosomici (aploidi) si riuniscono formando il nucleo (diploide) dello zigote (Fig. 9). Ogni specie vegetale ed animale conserva un numero fisso di cromosomi (per esempio nell'uomo ci sono $2n = 32$ cromosomi). Esistono tuttavia rare mutazioni (duplicazioni, traslocazioni, delezioni, fusioni cromosomiche) che possono modificare il cariotipo di un individuo. Le mutazioni cromosomiche

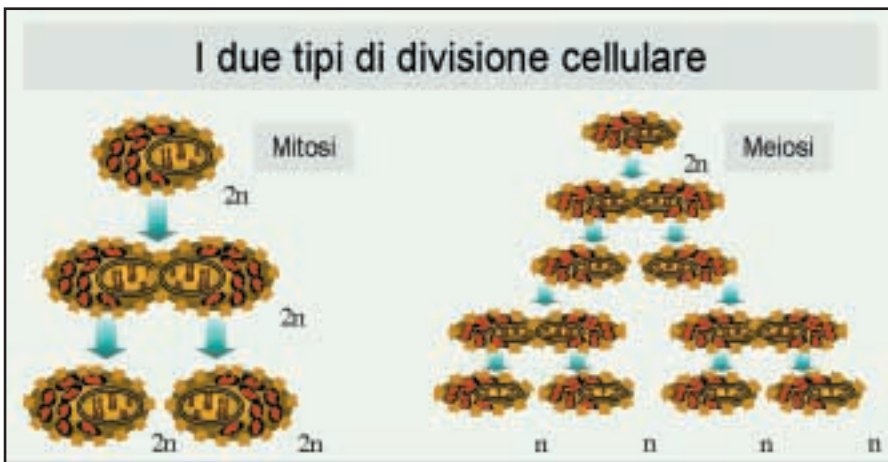


Figura 7 - La divisione delle cellule dei tessuti somatici, chiamata mitosi, non ha alcuna implicazione per la trasmissione ereditaria dell'informazione genetica alla generazione successiva. Ogni coppia di cellule originatisi da ogni divisione cellulare somatica, contiene esattamente lo stesso DNA della cellula parentale. Il corredo cromosomico di una cellula è detto "cariotipo diploide" ($2n$). Nel corso della meiosi si formano le cellule della riproduzione (spermatozoi e cellule-uovo). Il contenuto del DNA delle cellule somatiche diploidi viene dimezzato e la cellula diventa aploide (n). La riduzione meiotica del corredo cromosomico è essenziale per conservare inalterato nel corso della fecondazione il numero diploide di cromosomi che è tipico di ogni specie.

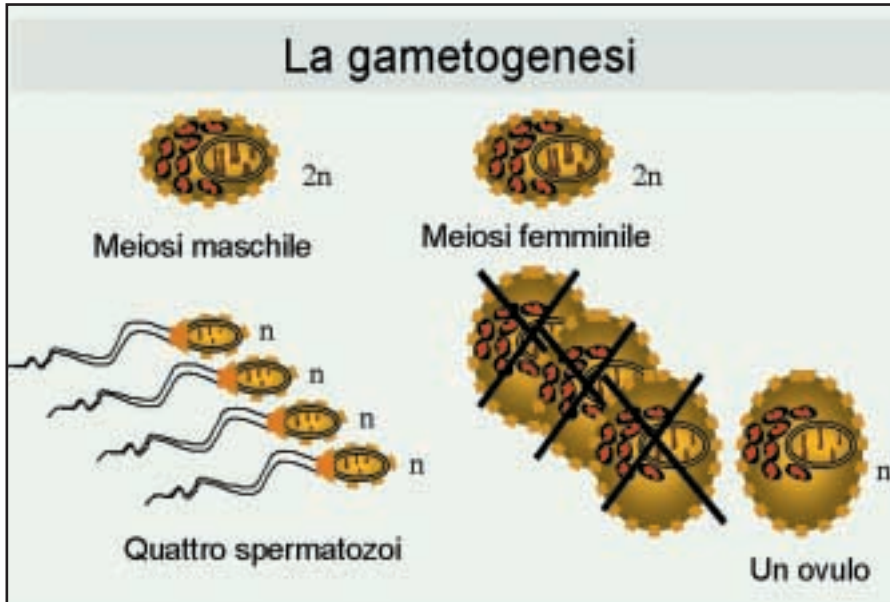


Figura 8 - Al termine di ogni ciclo di formazione dei gameti ciascuna cellula maschile diploide ($2n$) origina quattro spermatozoi aploidi (n). Ciascuna cellula femminile diploide origina quattro ovuli aploidi (n), tre dei quali degenerano e solo uno resta funzionale.

hanno spesso effetti deleteri o letali, e sono rapidamente eliminate dalle popolazioni tramite l'azione della selezione naturale. Nel corso della meiosi, i cromosomi di ogni coppia, di cui uno di origine materna e l'altro di origine paterna, sono appaiati e possono scambiarsi dei frammenti tramite il meccanismo del crossing-over. Il crossing-over produce ricombinazione (Fig. 10). La ricombinazione è un importante meccanismo di generazione della variabilità genetica, perché produce nuove sequenze di geni che originano dall'assortimento di segmenti di DNA ereditati in parte dalla madre ed in parte dal padre.

Il DNA mitocondriale (mtDNA) ha forma di una doppia elica circolare (Fig. 11), costituita da circa 15.000 - 20.000 nucleotidi nelle diverse specie. Il mtDNA viene replicato, indipendentemente dalle repliche del DNA nucleare, ogni qualvolta i mitocondri si dividono. Ogni cellula contiene da 5.000 a 10.000 mitocondri. Ogni mitocondrio contiene 10 o più molecole di mtDNA. Nel corso della gametogenesi il contenuto di citoplasma cambia significativamente, e quindi cambia il numero di mitocondri contenuti nei gameti. I mitocondri sono forniti interamente dall'uovo. Perciò, nel corso della fecondazione, è la cellula-

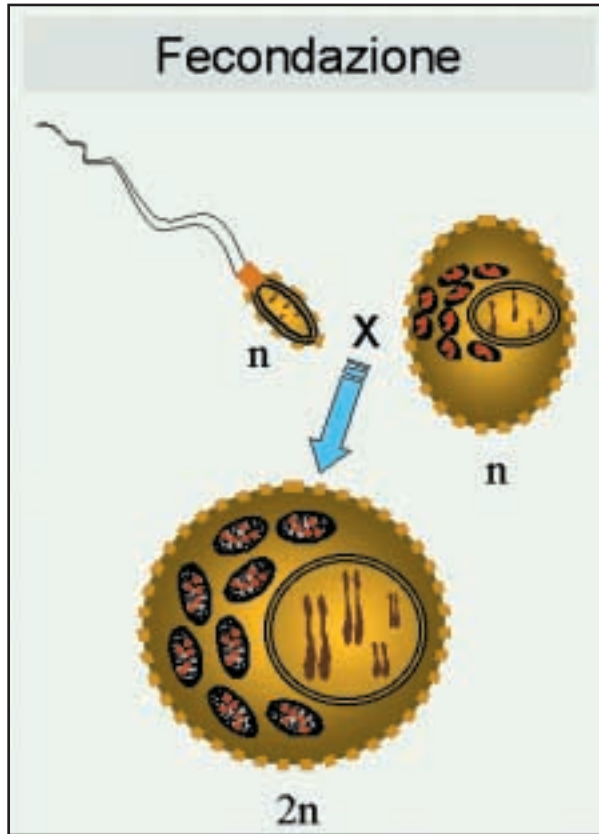


Figura 9 - Quando la cellula-uovo viene fecondata da uno spermatozoo, il nucleo cellulare materno e quello paterno si uniscono ed i due complementi cromosomici (aploidi) si riuniscono formando il nucleo (diploide) dello zigote.

uovo che trasmette tutti i mitocondri allo zigote. Il mtDNA è quindi trasmesso esclusivamente per via materna (Fig. 9). Il genoma mitocondriale è aploide e non ricombina. I diversi tipi di mtDNA che originano per mutazione e che sono presenti nelle popolazioni, sono chiamati “aplotipi mitocondriali”.

Il genoma dei vertebrati e di molti altri organismi viventi, è composto in gran parte da sequenze di DNA non codificanti, che apparentemente non hanno alcuna funzione (Fig. 12). Alcuni geni esistono in famiglie composte da gruppi di sequenze simili e che sono apparentemente deri-

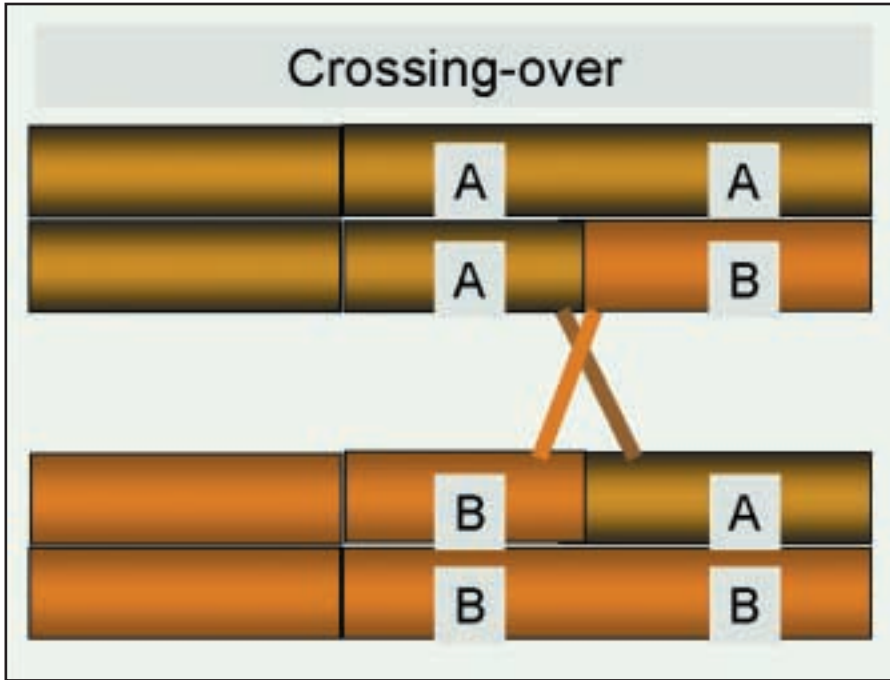


Figura 10 - Nel corso della meiosi, i cromosomi di ogni coppia, di cui uno di origine materna e l'altro di origine paterna, sono appaiati e possono scambiarsi dei frammenti tramite il meccanismo del crossing-over. Il crossing-over produce ricombinazione. La ricombinazione è un importante meccanismo di generazione della variabilità genetica, perchè produce nuove sequenze di geni che originano dall'assortimento di segmenti di DNA ereditati in parte dalla madre ed in parte dal padre.

vate le une dalle altre. I meccanismi che generano le famiglie di geni sono la duplicazione e la conversione genica. Le copie di un gene duplicato iniziano ad evolvere indipendentemente ed accumulano mutazioni differenti. L'effetto di queste mutazioni può essere duplice: il gene duplicato resta attivo ed acquisisce nuove funzioni (ad esempio, codifica per una nuova proteina), oppure viene inattivato da mutazioni che ne bloccano la funzionalità. In questo caso la copia diviene uno pseudogene. Ci sono altre famiglie di sequenze ripetute che sono state probabilmente generate da processi di trascrizione inversa. Nelle cellule sono presenti molecole di RNA che sono trascritte in DNA (tramite un enzima, analogo della DNA polimerasi, che si chiama polimerasi inversa) e, successivamente, sono inserite nei cromosomi. Questo DNA sembra essere d'origine esogena, ad esempio potrebbe derivare dalla trascrizione inversa di RNA

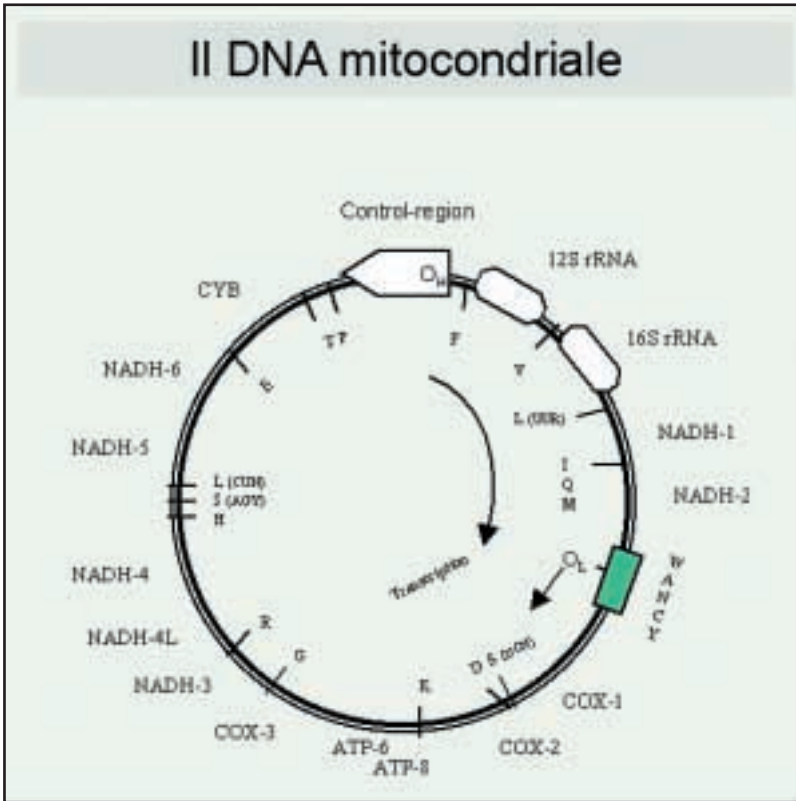


Figura 11 - Il DNA mitocondriale è una doppia elica circolare, costituita da circa 15.000 - 20.000 nucleotidi. Il mtDNA è replicato, indipendentemente dalle repliche delle cellule e del DNA nucleare, ogni qualvolta i mitocondri si dividono. Il genoma mitocondriale è aploide e non ricombina. I diversi tipi di mtDNA, che originano per mutazione e che sono presenti nelle popolazioni, sono chiamati "aplotipi mitocondriali". Lo spermatozoo è quasi completamente privo di citoplasma e quindi è privo di mitocondri. Viceversa, la cellula-uovo è ricca di citoplasma e di mitocondri (Fig. 8). Perciò, nel corso della fecondazione, la cellula-uovo trasmette tutti i mitocondri allo zigote. Il mtDNA è quindi trasmesso esclusivamente per via materna (Fig. 9).

virali. Una volta inserite nel genoma, queste sequenze evolvono per duplicazione genica. Attualmente non è chiaro se queste sequenze abbiano qualche funzione o se siano semplicemente costituite da DNA parassitario che, una volta inserito nel genoma, semplicemente si autopreserva duplicandosi incessantemente senza danneggiare il genoma dell'ospite. Tuttavia dati recenti dimostrano che certe sequenze ripetute possono avere funzioni, ad esempio come siti di regolazione del crossing-over e della ricombinazione.

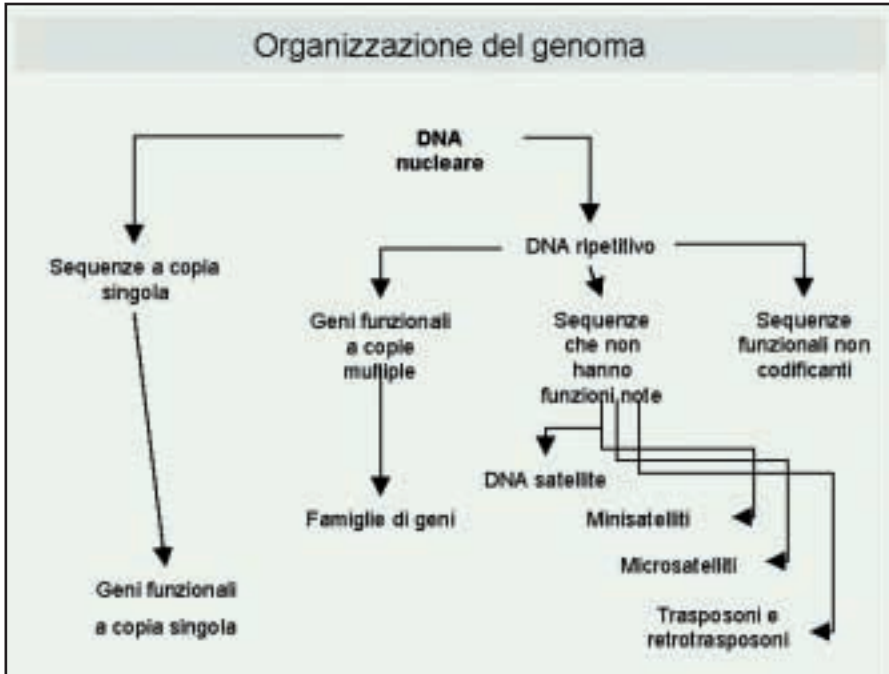


Figura 12 - Il genoma dei vertebrati è composto da sequenze codificanti a copia singola e, in gran parte, da sequenze di DNA non codificanti, che apparentemente non hanno alcuna funzione fisiologica. Alcuni geni esistono in famiglie composte da gruppi di sequenze simili e che sono apparentemente derivate le une dalle altre. Esistono sequenze non codificanti presenti in copie multiple nel genoma (DNA ripetitivo). Il DNA ripetitivo è classificato in tre gruppi principali: il DNA satellite, composto da unità ripetute molto lunghe che sono solitamente associate con i centromeri; DNA minisatellite (Fig. 14) e il DNA microsatellite (Fig. 15). Il DNA satellite di solito non è usato in genetica di popolazioni o in genetica forense, mentre il DNA minisatellite e microsatellite è molto utilizzato in genetica forense.

I geni, sequenze presenti in copia unica o in famiglie composte da un piccolo numero di copie dello stesso gene, costituiscono il DNA funzionale, non ripetitivo, e codificano per le proteine (Fig. 13). Le sequenze di DNA che compongono il gene sono organizzate in domini funzionali che hanno il ruolo di regolare la trascrizione: la prima parte del gene è costituita dal promotore, una sequenza di poche decine di nucleotidi cui si lega la RNA polimerasi. Seguono poi le sequenze codificanti (esoni) che di norma si alternano con tratti di sequenze che sono trascritte, ma non tradotte (introni). Il gene termina con le sequenze di terminazione, che interrompono la sintesi di RNA. I terminatori sono sequenze lunghe poche decine di nucleotidi. Il meccanismo della sintesi proteica si rea-

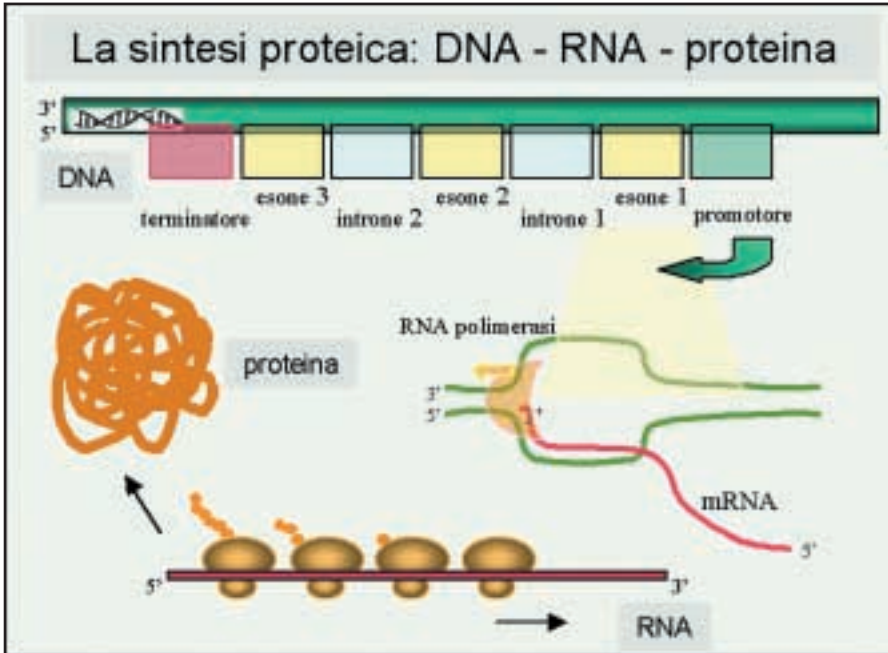


Figura 13 - Le sequenze di DNA che compongono il gene sono organizzate in domini funzionali che hanno il ruolo di regolare la trascrizione: la prima parte del gene è costituita dal promotore, una sequenza di poche decine di nucleotidi cui si lega la RNA polimerasi. Seguono poi le sequenze codificanti (esoni) che di norma si alternano con tratti di sequenze che sono trascritte, ma non tradotte (introni). Il gene termina con le sequenze di terminazione, che interrompono la sintesi di RNA. Il meccanismo della sintesi proteica si realizza tramite la trascrizione del DNA in RNA messaggero e la traduzione dell'RNA messaggero in proteina. La trascrizione si avvia quando il DNA di un gene viene denaturato, la doppia elica si apre in prossimità del promotore ed uno dei due filamenti singoli agisce come stampo per la sintesi dell'RNA. DNA ed RNA hanno struttura molecolare del tutto simile, cioè sono composti entrambi da sequenze di quattro nucleotidi, ma, nell'RNA la timina (T) è sostituita dall'uracile (U). I ribonucleotidi (U, A, C, G) presenti nel citoplasma vengono assemblati linearmente tramite l'azione enzimatica dell'RNA polimerasi. Al termine della trascrizione l'RNA messaggero è composto da una sequenza complementare agli esoni ed introni del gene. Questo RNA primario viene processato e tutti gli introni sono rimossi (RNA maturo). La sequenza di un RNA maturo viene tradotta in sequenza proteica. Le proteine sono composte da aminoacidi, che sono assemblati linearmente nel corso della traduzione grazie al meccanismo del codice genetico. Il codice genetico è definito dalle triplette di nucleotidi dell'RNA maturo.

lizza tramite due passaggi: la trascrizione del DNA in RNA messaggero e la traduzione dell'RNA messaggero in proteina. La trascrizione si avvia quando il DNA di un gene viene denaturato, la doppia elica si apre in prossimità del promotore ed uno dei due filamenti singoli agisce come stampo per la sintesi dell'RNA. DNA ed RNA hanno struttura moleco-

lare del tutto simile, cioè sono composti entrambi da sequenze di quattro nucleotidi, ma, nell'RNA la timina (T) è sostituita dall'uracile (U). I ribonucleotidi (U, A, C, G) presenti nel citoplasma vengono assemblati linearmente tramite l'azione enzimatica dell'RNA polimerasi. Al termine della trascrizione l'RNA messaggero è composto da una sequenza complementare agli esoni ed introni del gene. Questo RNA primario viene processato e tutti gli introni sono rimossi (RNA maturo). La sequenza di un RNA maturo viene tradotta in sequenza proteica. Le proteine sono composte da aminoacidi, che sono assemblati linearmente nel corso della traduzione grazie al meccanismo del codice genetico. Il codice genetico è definito dalle triplette di nucleotidi dell'RNA maturo.

Esistono poi sequenze non codificanti presenti in copie multiple nel genoma di ogni specie (DNA ripetitivo). Il DNA ripetitivo è classificato in tre gruppi principali:

- DNA satellite: è composto da unità ripetute molto lunghe (fino a 5.000.000 di nucleotidi) che sono solitamente associate con i centromeri (le zone alle quali si agganciano le fibre del fuso mitotico e che controllano la ripartizione dei cromosomi nelle due cellule figlie nel corso di ogni divisione somatica o gametica). Il DNA satellite non è usato in genetica di popolazioni o in genetica forense.
- DNA minisatellite (Fig. 14): è presente in centinaia o migliaia di loci nei genomi eucarioti. Le sequenze ripetute dei minisatelliti sono composte da unità lunghe più di 10 nucleotidi, e sono presenti in copie multiple che producono blocchi di 500 - 30.000 nucleotidi. Alcuni minisatelliti sono ipervariabili e sono ampiamente utilizzati in genetica forense per ottenere i DNA fingerprinting. Questi loci sono tipizzati usando sonde multilocus (multi-locus probes - MLP). Tramite analisi molecolare sono stati identificati alcuni dei loci che compongono i minisatelliti (loci VNTR). Questi loci possono essere individuati e tipizzati tramite l'uso di sonde specifiche (single-locus probes - SLP).
- DNA microsatellite (Fig. 15): è presente in molte migliaia di loci nei genomi eucarioti. I microsatelliti sono costituiti da sequenze molto corte (da 2 a 8 nucleotidi) che sono ripetute poche volte e producono blocchi di poche decine o poche centinaia di nucleotidi ad ogni locus. I microsatelliti sono ampiamente utilizzati in genetica forense e sono tipizzati tramite PCR.

Le diverse categorie di sequenze di DNA funzionale o non-funzionale evolvono con meccanismi mutazionali differenti, che sono in relazione con la struttura e funzione del DNA.

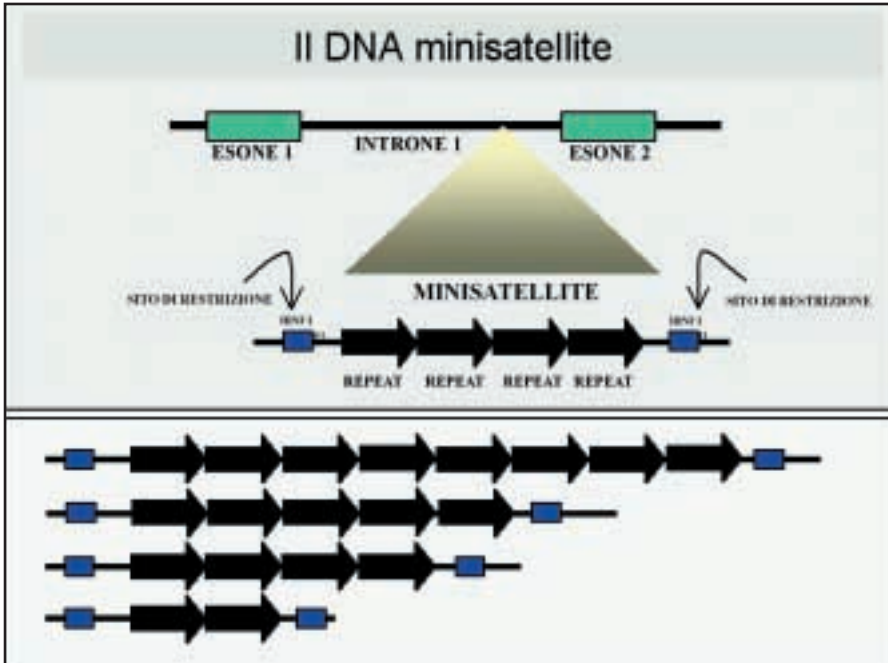


Figura 14 - Il DNA minisatellite è presente in centinaia o migliaia di loci nei genomi eucarioti. Le sequenze ripetute dei minisatelliti sono composte da unità lunghe più di 10 nucleotidi, e sono presenti in copie multiple che producono blocchi di 500 - 30.000 nucleotidi. Alcuni minisatelliti sono ipervariabili e sono ampiamente utilizzati in genetica forense per ottenere i DNA fingerprinting. Questi loci sono tipizzati usando sonde multilocus (multi-locus probes; MLP). La parte inferiore della figura indica schematicamente la composizione di quattro diversi alleli ad un locus minisatellite.

- Sostituzioni nucleotidiche ed aminoacidiche (Fig. 16). Il tipo di mutazione più semplice è la sostituzione di un nucleotide con un altro ad un punto di una catena di DNA. Le sostituzioni nucleotidiche sono chiamate anche mutazioni puntiformi. Una mutazione puntiforme può avvenire al di fuori dei geni, cioè delle sequenze di DNA che codificano per le proteine. Le mutazioni che non cambiano le sequenze aminoacidiche sono dette mutazioni silenti (sinonime). Le mutazioni che modificano il codice genetico e provocano una sostituzione aminoacidica sono dette non-sinonime.
- Inserzioni e delezioni di singoli o di serie di nucleotidi (Fig. 16). Queste mutazioni possono modificare la lettura del codice genetico oppure inattivare il gene.

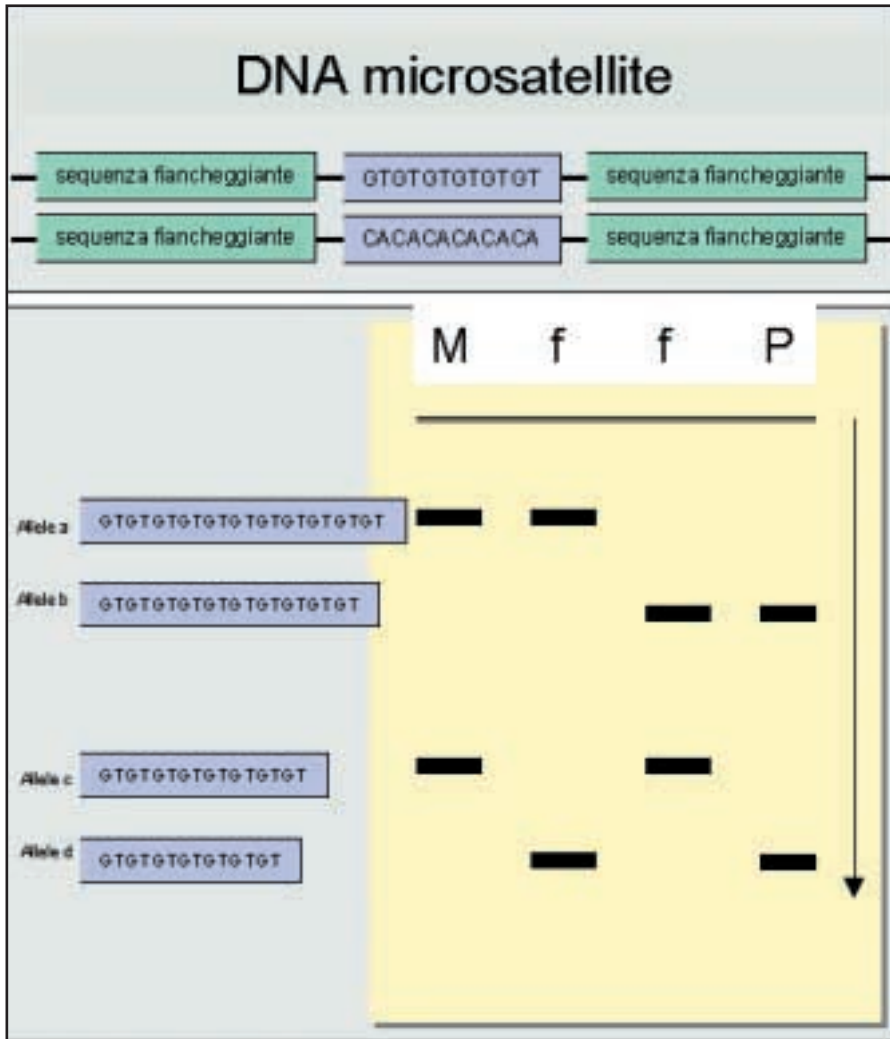


Figura 15 - DNA microsatellite. I microsatelliti sono costituiti da sequenze molto corte, cioè da 2 a 8 nucleotidi che sono ripetuti poche volte e producono blocchi di poche decine o poche centinaia di nucleotidi ad ogni locus. I microsatelliti sono ampiamente utilizzati in genetica forense e sono tipizzati tramite PCR. La parte inferiore della figura indica schematicamente la composizione di quattro diversi alleli ad un locus minisatellite, ed il risultato di una separazione elettroforica di questi alleli.

- Crossing-over e ricombinazione. Il crossing-over può essere simmetrico (Fig. 10) o asimmetrico (Fig. 16). Il crossing-over simmetrico produce scambi di sequenze corrispondenti fra due cromosomi e produce

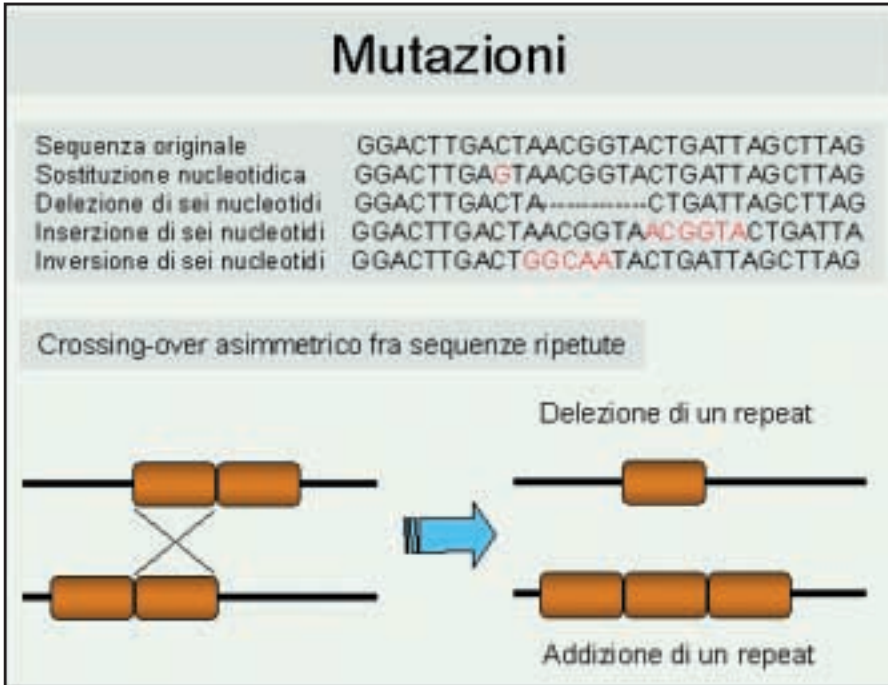


Figura 16 - Mutazioni. Il tipo di mutazione più semplice è la sostituzione di un nucleotide con un altro ad un punto della catena del DNA (mutazioni puntiformi). Le inserzioni e delezioni di singoli o di serie di nucleotidi possono modificare la lettura del codice genetico oppure inattivare il gene. Crossing-over e ricombinazione. Il crossing-over può essere simmetrico (Fig. 10) o asimmetrico. Il crossing-over asimmetrico avviene più frequentemente fra sequenze di DNA satellite o minisatellite, cioè fra sequenze di DNA ripetitivo che possono allinearsi imprecisamente. Il crossing-over asimmetrico origina la delezione di un frammento di DNA da un cromatidio e la sua inserzione nell'altro cromatidio. Il crossing-over asimmetrico può avvenire fra i due cromatidi dello stesso cromosoma o fra due cromosomi diversi.

ricombinazione genetica (Fig. 17). Il crossing-over asimmetrico avviene più frequentemente fra sequenze di DNA satellite o minisatellite, cioè fra sequenze di DNA ripetitivo che possono allinearsi non precisamente. Il crossing-over asimmetrico origina la delezione di un frammento di DNA da un cromatidio e la sua inserzione nell'altro cromatidio. Il crossing-over asimmetrico può avvenire fra i due cromatidi dello stesso cromosoma o fra due cromosomi diversi.

- DNA slippage (Fig. 18). Lo slippage avviene durante la replicazione quando il DNA nascente si dissocia e riassocia temporaneamente dal DNA stampo. Nel corso della replicazione di sequenze non ripetitive

l'eventuale dissociazione del cromatidio figlio di solito non genera mutazioni, poiché il DNA nascente si può riassociare solo ed esattamente nel punto complementare del DNA stampo. Nel corso della replicazione di DNA ripetitivo, il DNA nascente dissociato può riassociarsi in un altro punto del DNA stampo. Quando la replicazione continua, il DNA nascente risulterà più lungo dello stampo.

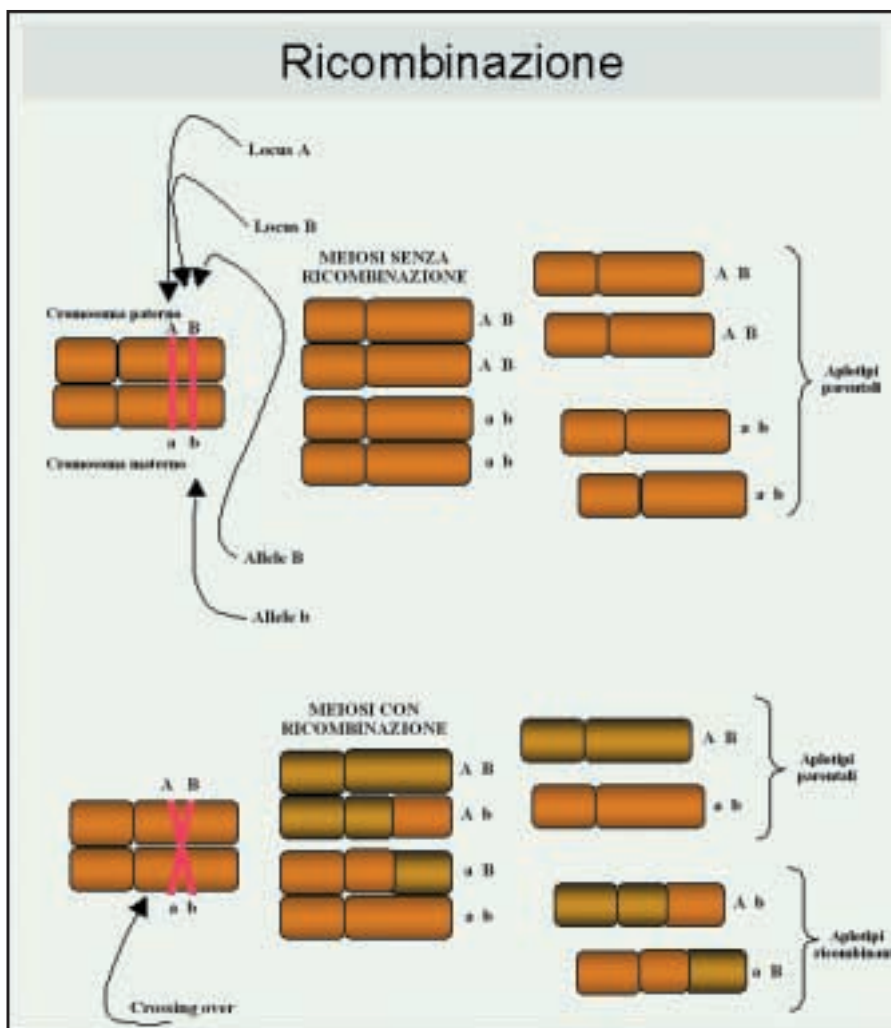


Figura 17 - Il crossing-over simmetrico produce scambi di sequenze corrispondenti fra due cromosomi e produce ricombinazione genetica.

- Conversione genica (Fig. 19): la conversione genica produce trasferimento di una sequenza di DNA da un allele ad un altro.

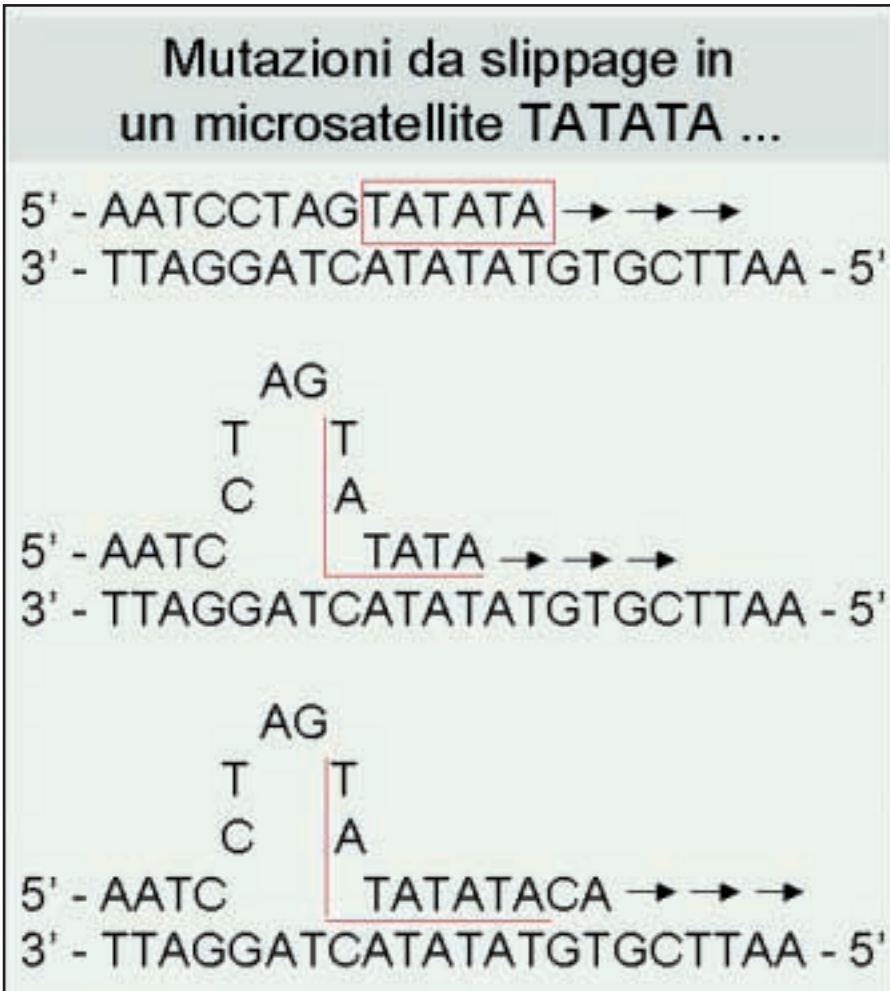


Figura 18 - Il DNA slippage avviene durante la replicazione quando il DNA nascente si dissocia e riassocia temporaneamente dal DNA stampo. Nel corso della replicazione di sequenze non ripetitive l'eventuale dissociazione del cromatidio figlio di solito non genera mutazioni, poiché il DNA nascente si può riassociare solo ed esattamente nel punto complementare del DNA stampo. Nel corso della replicazione di DNA ripetitivo, il DNA nascente dissociato può riassociarsi in un altro punto del DNA stampo. Quando la replicazione continua, il DNA nascente risulta più lungo dello stampo.

Mutazioni e polimorfismi genetici

Le mutazioni generano variabilità genetica negli individui e nelle popolazioni. Un gene variabile è definito polimorfico. I polimorfismi indicano la presenza di due o più varianti di una sequenza di DNA. Ovviamente, i polimorfismi dei geni codificanti possono generare polimorfismi proteici (ad esempio, alloenzimi, gruppi sanguigni, immunoglobine, ecc.), oltre che polimorfismi dei caratteri fenotipici (colore degli occhi, della pelle, struttura dei capelli, impronte digitali, ecc.). Tutti questi caratteri possono essere utilizzati come marcatori per l'identificazione ed individuazione dei campioni nella scienza forense. Le sequenze ipervariabili di DNA non codificante, che apparentemente non sono sottoposte a forti pressioni di selezione naturale e quindi evolvono rapidamente e neutralmente, costituiscono i marcatori genetici più utili ed affidabili per acquisire evidenze in genetica forense.

Mutazioni nei minisatelliti. I minisatelliti sono molto variabili, con tassi di mutazione nell'ordine di 10^{-3} per frammento per gamete. Ogni allele muta una volta ogni circa mille cicli di gametogenesi, il che vuol dire che ci si aspetta di trovare una mutazione ad ogni gametogenesi analizzando circa mille alleli indipendenti in un profilo multilocus. Questi elevati tassi di mutazione generano il gran numero di alleli che è necessario per l'individuazione, ma possono anche generare frammenti spuri

di difficile assegnazione. Per esempio, eventuali mutazioni somatiche possono generare DNA fingerprinting diversi in campioni di DNA estratti da differenti tessuti dello stesso individuo. In questo caso, l'identificazione potrebbe essere problematica. Inoltre, le mutazioni gametiche possono generare differenze fra i genitori ed i figli. In entrambi i casi queste mutazioni potrebbero generare falsi negativi e quindi produrre false diagnosi di esclusione. Tuttavia, nell'arco di tempo di una generazione queste mutazioni restano pur sempre rare ed in pratica non dovrebbero interferire sui risultati delle analisi gene-

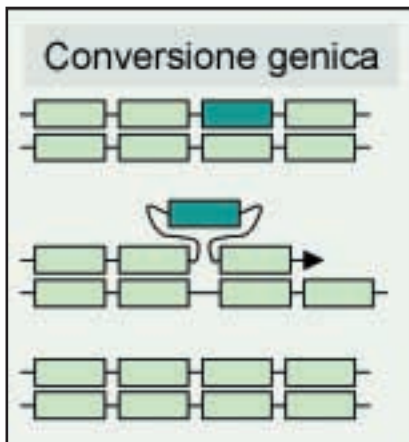


Figura 19 - La conversione genica produce il trasferimento di una sequenza di DNA da un allele ad un altro. Un allele viene così "convertito" in un altro.

tiche. Un test di parentela si basa sull'analisi dei profili multilocus trasmessi dai genitori ai figli nell'ambito di una generazione. Con tassi di mutazione dell'ordine di 10^{-3} per frammento per gamete, ci sarebbero probabilità di trovare una mutazione se il profilo materno ed il profilo paterno fossero composti in totale da mille alleli diagnostici. Normalmente nelle diagnosi di parentela si utilizzano circa 20 - 40 frammenti per ogni coppia di genitori, quindi le probabilità di individuare una mutazione restano piuttosto basse. Una mutazione somatica o gametica dovrebbe modificare il profilo di un singolo frammento o di un singolo allele in un sistema multilocus. Per esempio, se, sulla base dei dati ottenuti da un singolo locus, un'allele che non fosse presente nel padre presunto apparisse a seguito di una mutazione nel profilo del figlio, quel padre potrebbe essere non correttamente escluso. E' molto improbabile che altre mutazioni abbiano contemporaneamente modificato i profili ad altri loci indipendenti, o in profili multilocus ottenuti utilizzando altri enzimi di restrizione. Una singola mutazione non può pertanto essere usata come prova di esclusione. Il significato di una singola mutazione deve essere valutato analizzando diversi loci in sistemi a locus singolo, od utilizzando due o più enzimi di restrizione in sistemi multilocus.

Mutazioni nei microsatelliti. I microsatelliti sono sequenze composte da un motivo semplice di 2 - 8 nucleotidi, che è ripetuto in tandem un certo numero di volte, con o senza interruzioni dovute all'inserimento di altri nucleotidi o di altre sequenze. I microsatelliti presentano elevati livelli di polimorfismo. I microsatelliti sono stati individuati nel genoma di tutti gli organismi fino ad ora analizzati e sono distribuiti, in modo più o meno casuale, nei cromosomi. Di solito non sono presenti all'interno delle sequenze codificanti dei geni (esoni), mentre possono essere presenti negli introni. La composizione delle sequenze dei microsatelliti è variabile: i motivi poli(A)/poli(T) sono molto comuni nei vertebrati, ma non sono utilizzabili come marcatori genetici perché sono estremamente instabili durante la PCR. I motivi CA/GT sono fra i più comuni dinucleotidi. Altri dinucleotidi sono AT/TA e AG/TC. Esistono poi microsatelliti composti da sequenze ripetute di trinucleotidi (esempio CAG, o AAT) o di tetranucleotidi. In alcuni casi le sequenze fiancheggianti sono conservate nel corso dell'evoluzione. E' possibile utilizzare primer di PCR conservati per amplificare ed analizzare microsatelliti in specie diverse. Le mutazioni che determinano aggiunta o perdita di una o più unità dei repeat sono molto più frequenti delle sostituzioni nucleotidiche. Le stime dei tassi di mutazione nei microsatelliti dei vertebrati sono di 10^{-4} - 10^{-5} mutazioni per locus per ogni generazione. Queste mutazioni sono

quindi di uno o due ordini di grandezza minori dei tassi di mutazione dei minisatelliti. I meccanismi mutazionali che determinano la variazione del numero dei repeats e quindi la variazione del peso molecolare degli alleli ai loci microsatellite sono lo slippage ed il crossing-over asimmetrico. Alcuni risultati sperimentali suggeriscono che lo slippage è probabilmente il principale meccanismo responsabile di mutazione dei microsatelliti.

Sostituzioni nucleotidiche. Le sequenze di DNA degli esoni sono conservate dalla selezione naturale, che elimina tutte quelle mutazioni che producono proteine malfunzionanti o che impediscono la sintesi proteica. Tuttavia il codice genetico è ridondante, cioè esistono più triplette nucleotidiche (codoni) che codificano per lo stesso aminoacido. La ridondanza è determinata particolarmente dal nucleotide in terza posizione in ogni codone. Pertanto, molte sostituzioni nucleotidiche che avvengono alle terze posizioni dei codoni sono sinonime. Le mutazioni sinonime sono molto più frequenti delle sostituzioni nucleotidiche non-sinonime. Le sostituzioni nucleotidiche ed i riarrangiamenti (inserzioni e delezioni) sono molto più frequenti nelle sequenze di DNA non codificanti e di DNA ripetitivo. In particolare, sono molto variabili le sequenze di DNA che costituiscono le regioni di controllo della replicazione dell'mtDNA. Sequenze nucleotidiche di DNA non-codificante, di introni e soprattutto della regione di controllo dell'mtDNA, rivelano ampia variabilità nelle popolazioni e sono usate in genetica forense.

VARIABILITÀ GENETICA NEGLI INDIVIDUI E NELLE POPOLAZIONI

I meccanismi dell'eredità: le leggi di Mendel

La genetica moderna nasce grazie agli studi di Gregor Mendel pubblicati nel 1866. Nei suoi esperimenti Mendel utilizzò linee pure di piante di pisello che presentavano caratteri fenotipici ben identificabili e contrastanti. Per esempio, alcune linee presentavano sempre il seme di colore giallo, mentre in altre il colore del seme era verde. Mendel eseguì diversi schemi di incrocio fra linee pure, descrisse le frequenze dei caratteri fenotipici che apparivano in successive generazioni di incrocio e reincrocio, e sviluppò un modello genetico che potesse spiegare i risultati degli incroci. Lo scopo degli esperimenti di Mendel fu la determinazione delle leggi che controllano la trasmissione ereditaria dei "caratteri" fenotipici. Mendel

ipotizzò che l'espressione fenotipica di ciascun carattere fosse determinata da "fattori genetici" discreti, che in seguito saranno chiamati "geni", che sono trasmessi integri nel corso delle generazioni dai genitori ai figli. Per esempio, dall'incrocio fra due linee parentali pure di piselli, una con seme verde e l'altra con seme giallo, Mendel ottenne una prima generazione (F1) che mostrava il 100% dei semi gialli, per effetto della dominanza del carattere giallo sul carattere verde. Incrociando fra di loro le piante F1, Mendel ottenne una successiva generazione (F2) in cui riapparve il carattere seme verde, che apparentemente era scomparso nella F1, con una frequenza che è esattamente prevedibile se applichiamo il modello di eredità mendeliana che è descritto nella figura 20.

La prima legge di Mendel è detta "legge della segregazione indipendente degli alleli", e stabilisce che, nel corso della meiosi, i due alleli sono separati (segregati) indipendentemente in cellule aploidi diverse. La seconda legge di Mendel è detta "legge dell'assortimento indipendente", e stabilisce che alleli a loci diversi, collocati su cromosomi diversi, sono associati (assortiti) indipendentemente nel corso della meiosi. Le leggi di Mendel consentono di prevedere le proporzioni dei diversi genotipi che sono prodotte ad ogni generazione, come semplice e diretta conseguenza della segregazione e dell'assortimento indipendente degli alleli ad uno o a più loci (Fig. 21).

Oggi sappiamo che ogni individuo eredita uno dei cromosomi di ogni coppia dalla madre ed uno dal padre. Ogni gene è collocato in un particolare punto del cromosoma ("locus", plurale "loci"), ed è presente in duplice copia, ognuna delle quali è chiamata "allele" (Fig. 4). La localizzazione dei loci genici nei cromosomi consente di ricostruire la mappa cromosomica. L'identificazione degli alleli presenti ai loci polimorfici consente di identificare i genotipi individuali e di stimare la variabilità genetica nelle popolazioni. I due alleli ad un locus possono avere caratteristiche genetiche identiche (il locus è "omozigote"), oppure differenti (il locus è "eterozigote"; Fig. 17). In questo caso il locus è polimorfico. Alternativamente, il locus è detto monomorfico. Ogni individuo diploide può avere, al massimo, due diversi alleli ad ogni locus, ma il locus può avere alleli multipli che sono distribuiti negli individui della popolazione. I geni funzionali esprimono o contribuiscono ad esprimere i caratteri fenotipici (proteine, tratti fisiologici, morfologici o comportamentali). Perciò i polimorfismi genetici possono esprimere polimorfismi fenotipici e quindi determinano variabilità fenotipica fra gli individui che compongono una popolazione o fra individui che appartengono a popolazioni diverse. Per esempio, consideriamo il gene A che ha due alleli a_1 e a_2 .

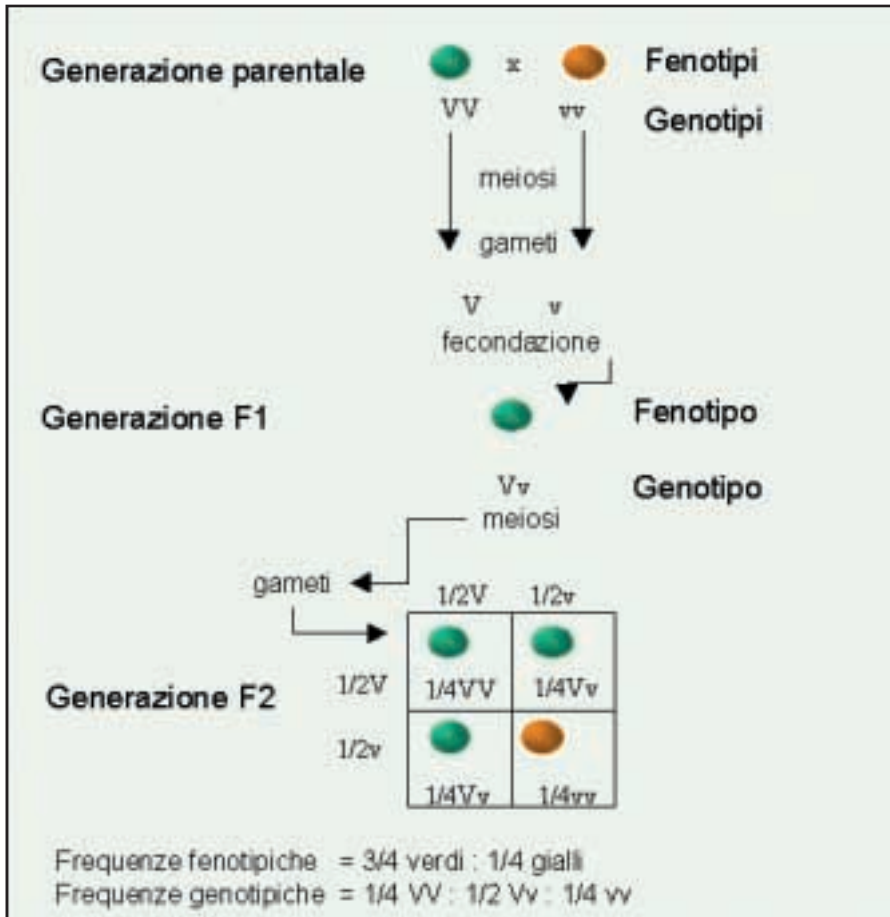


Figura 20 - Schema di incrocio mendeliano: incrocio fra due linee parentali pure di piselli, una con seme verde e l'altra con seme giallo. La prima generazione (F1) mostra il 100% dei semi gialli, per effetto della dominanza del carattere giallo sul carattere verde. Incrociando fra di loro le piante F1, Mendel ottenne una successiva generazione (F2) in cui riapparve il carattere seme verde, che apparentemente era scomparso nella F1, con una frequenza che è esattamente prevedibile se applichiamo il modello di eredità mendeliana.

In alcuni individui della popolazione i due alleli sono identici, allora il genotipo è omozigote (a_1a_1 oppure a_2a_2). Negli altri individui i due alleli sono differenti, allora il genotipo è eterozigote (a_1a_2 , oppure l'equivalente a_2a_1). Ad ogni particolare ciclo riproduttivo, ogni individuo della generazione parentale genera gameti che, ad ogni locus, hanno alleli identici o diversi, a seconda che l'individuo sia omozigote o eterozigote. Ogni figlio

Proporzioni genotipiche attese per geni a segregazione indipendente					
Genotipi dei:					
Genitori	a/a x a/a	A/A x A/A	A/A x a/a	A/a x a/a	A/a x A/a
Gameti	a a	A A	A a	A a a A/a a/a	A a A AA Aa a aA aa
Figli	a/a	A/A	A/a		

Figura 21 - La prima legge di Mendel è detta “legge della segregazione indipendente degli alleli”, e stabilisce che, nel corso della meiosi, i due alleli sono separati (segregati) indipendentemente in cellule aploidi diverse. La seconda legge di Mendel è detta “legge dell’assortimento indipendente”, e stabilisce che alleli a loci diversi, collocati su cromosomi diversi, sono associati (assortiti) indipendentemente nel corso della meiosi. Le leggi di Mendel consentono di prevedere le proporzioni dei diversi genotipi che sono prodotte ad ogni generazione, come semplice e diretta conseguenza della segregazione e dell’assortimento indipendente degli alleli ad uno o a più loci.

può ricevere un solo allele dal padre ed uno solo dalla madre. La scelta degli alleli paterni e materni che vengono trasmessi alla prole è “casuale”, nel senso che entrambi gli alleli hanno le medesime probabilità di essere trasmessi alla prole. Per esempio, gli individui con genotipo omozigote a_1a_1 generano solo gameti a_1 e possono trasmettere alla prole solo l’allele a_1 , mentre gli individui eterozigoti (per esempio a_1a_2) generano gameti sia a_1 che a_2 , e possono trasmettere sia l’allele a_1 che l’allele a_2 .

Le leggi di Mendel consentono di stimare le frequenze dei genotipi anche in popolazioni che non sono composte da linee pure (ciascuna delle quali è costituita da individui identici per il carattere considerato), ma da individui che presentano l’una o l’altra delle due forme e che sono presenti nella popolazione in determinate frequenze. Per esempio, in una popolazione, composta da individui che si riproducono casualmente, in cui al locus A (con due alleli a_1 e a_2) l’allele a_1 sia presente con frequenza $p = 0.60$ (l’allele a_1 è presente nel 60% degli individui e quindi nel 60% dei gameti), la frequenza del genotipo omozigote a_1a_1 sarà semplicemente $0.60 \times 0.60 = 0.36$. Ovviamente, la frequenza dell’allele a_2 sarà $q = 1 - p = 0.40$. Quindi, la frequenza del genotipo omozigote a_2a_2 sarà $0.40 \times 0.40 = 0.16$. La frequenza dei genotipi eterozigoti a_1a_2 sarà $0.60 \times 0.40 = 0.24$, uguale alla frequenza degli eterozigoti a_2a_1 . Quindi la frequenza

totale degli eterozigoti sarà $2 \times 0.60 \times 0.40 = 0.48$. Le proporzioni dei genotipi al locus A saranno pertanto $= p^2 + 2pq + q^2 = 0.36 + 0.48 + 0.16 = 1.00$.

I genotipi e le frequenze genotipiche attese sulla base delle leggi di Mendel sono ancora oggi il fondamento dei test di parentela. Per esempio, se una madre con genotipo a_1a_2 ha un figlio con genotipo a_1a_3 , allora è necessario che l'allele a_3 sia stato trasmesso dal padre. In questo caso, eventuali padri presunti con genotipi a_1a_1 , a_1a_2 o a_2a_2 , possono essere esclusi come padri biologici di quel figlio. Un padre presunto che abbia l'allele a_3 ha una certa probabilità di essere il padre biologico. Questa probabilità può essere stimata quantitativamente utilizzando informazioni sulle frequenze dell'allele a_3 nella popolazione ed usando le appropriate procedure statistiche.

I meccanismi dell'eredità: associazione fra geni ("linkage")

I cromosomi sono trasmessi dai genitori ai figli come unità integre, perciò i geni che sono collocati su cromosomi diversi sono ereditati indipendentemente gli uni dagli altri (assortimento indipendente). Al contrario, i geni che mappano vicini sullo stesso cromosoma tendono ad essere ereditati assieme e quindi sono detti linked (legati). Tuttavia, nel corso della meiosi, il crossing-over produce ricombinazione, la quale rompe i gruppi di linkage (Fig. 17). Quanto più due geni sono fisicamente vicini su un cromosoma, tanto più saranno ereditati come una singola unità. Il linkage è una delle principali eccezioni alle leggi di Mendel. Quando due loci sono in equilibrio di linkage (LE), le frequenze di tutte le possibili combinazioni alleliche dipendono solo dalle frequenze degli alleli nella popolazione. Le frequenze delle combinazioni alleliche si ottengono calcolando i prodotti delle frequenze di tutte le possibili coppie di alleli. Se certe combinazioni alleliche sono più frequenti dell'atteso, allora il locus è in disequilibrio di linkage (LD).

Consideriamo due loci: A (con due alleli a_1 e a_2) e B (con due alleli b_1 e b_2). I due loci mappano sullo stesso cromosoma, ma non sono vicini. Ogni individuo può avere uno dei seguenti genotipi al locus A : a_1a_1 , a_1a_2 o a_2a_2 ; ed al locus B : b_1b_1 , b_1b_2 o b_2b_2 . Considerando i due loci assieme, è possibile generare nove diversi genotipi:

$a_1a_1b_1b_1$
 $\quad b_1b_2$
 $\quad b_2b_2$
 $a_1a_2b_1b_1$

$$\begin{array}{c}
 b_1b_2 \\
 b_2b_2 \\
 a_2a_2b_1b_1 \\
 b_1b_2 \\
 b_2b_2
 \end{array}$$

Questi genotipi possono generare quattro possibili tipi di gameti (aplotipi):

$$\begin{array}{c}
 a_1b_1 \\
 a_1b_2 \\
 a_2b_1 \\
 a_2b_2
 \end{array}$$

Gli individui che sono omozigoti ad entrambi i loci (doppi omozigoti) possono trasmettere solo un tipo di gamete (per esempio, $a_1a_1b_1b_1$ può trasmettere solo l'aplotipo a_1b_1). Gli individui omozigoti ad un solo locus possono trasmettere due tipi di gameti (per esempio, $a_1a_1b_1b_2$ può trasmettere gli aplotipi a_1b_1 e a_1b_2). I doppi eterozigoti $a_1a_2b_1b_2$ possono trasmettere solo due tipi di gameti, cioè i due aplotipi parentali, per esempio a_1b_1 , e a_2b_2 , in assenza di ricombinazione, oppure quattro tipi di gameti, e cioè i due aplotipi parentali più i due aplotipi ricombinanti, a_1b_2 , a_2b_1 e a_2b_2 , in presenza di ricombinazione. Se la probabilità di ricombinazione è c , ogni gamete ricombinante è trasmesso con probabilità $c/2$ ed ogni gamete parentale con probabilità $(1 - c)/2$. Tutti i loci che mappano su diversi cromosomi non sono associati, segregano indipendentemente e quindi trasmettono i quattro aplotipi gametici con eguale probabilità. I doppi eterozigoti a questi loci producono i quattro tipi di gameti con eguale probabilità, corrispondente a $c = 0.5$. Perciò, i valori di c sono compresi fra un minimo di $c = 0.0$ (per coppie di loci contigui sullo stesso cromosoma, con probabilità di ricombinazione nulla) e $c = 0.5$ (per coppie di loci che mappano su cromosomi diversi e che segregano in modo completamente indipendente). Il disequilibrio di linkage (LD) indica che la probabilità con cui un individuo eredita un allele al locus A , dipende anche dalla probabilità con cui eredita un allele al locus B . Il coefficiente di LD fra gli alleli a_1 e b_1 è dato dalla differenza fra la frequenza osservata dell'aplotipo a_1b_1 , $p(a_1b_1)$, e la sua frequenza attesa, che corrisponde al prodotto delle due frequenze alleliche $p(a_1)$ e $p(b_1)$:

$$LD = p(a_1b_1) - p(a_1)p(b_1)$$

Per esempio, se le frequenze alleliche sono:

$$a_1 = 0.9$$

$$a_2 = 0.1$$

$$b_1 = 0.6$$

$$b_2 = 0.4$$

le frequenze attese delle quattro combinazioni gametiche sono:

$$a_1b_1 = 0.54$$

$$a_1b_2 = 0.36$$

$$a_2b_1 = 0.06$$

$$a_2b_2 = 0.04$$

Se le frequenze osservate di questi quattro aplotipi sono diverse, e certe combinazioni sono più frequenti dell'atteso, allora è possibile che esista disequilibrio di linkage. La significatività delle discrepanze tra frequenze osservate ed attese può essere determinata tramite gli appropriati test statistici, fra cui il test del chi quadrato.

In genetica forense è importante utilizzare loci indipendenti che non siano in LD nella popolazione di riferimento. Infatti, le procedure basilari per ricavare le frequenze dei genotipi multilocus a partire dalle frequenze degli alleli ai singoli loci si basano sulla regola del prodotto che richiede l'indipendenza dei fattori (i loci).

Il LD può essere prodotto da mutazioni recenti che non sono ancora state randomizzate tramite ricombinazione nel genoma, oppure dalla selezione naturale, che favorisce la permanenza di certe combinazioni alleliche che hanno un significato funzionale ed adattativo. Una popolazione oggetto di immigrazione, oppure che si è originata dal mischiarsi di due sottopopolazioni geneticamente differenziate, può essere in disequilibrio di linkage, anche se entrambe le sottopopolazioni sono equilibrio di linkage. Il valore di LD in popolazioni miste è proporzionale alle differenze fra le frequenze alleliche delle due sottopopolazioni. Si può dimostrare che:

$$LD = mI mII (p(a_1)I - p(a_1)II)(p(b_1)I - p(b_1)II)$$

Con $p(a_1)$ e $p(b_1)$ = frequenze alleliche nelle sottopopolazioni I e II; mI e mII = proporzioni delle due sottopopolazioni nella popolazione totale.

Il LD viene, in teoria, eliminato rapidamente tramite la ricombinazione. Il valore di LD decade nel corso delle generazioni, con un tasso di decadimento da una generazione alla successiva che corrisponde a:

$$LD' = (1 - c)LD$$

Il decadimento di LD è massimo se $c = 0.5$, cioè se i geni sono non-

linked. In caso di popolazioni miste il LD decade a $LD' = (1 - 0.5)LD = 0.5LD$, in altre parole si dimezza dopo la prima generazione di riproduzione casuale fra le sottopopolazioni.

I geni nelle popolazioni

Lo scopo della genetica delle popolazioni è descrivere la composizione genetica delle popolazioni, e di capire le cause che ne determinano il cambiamento (forze evolutive). Ogni specie è composta da una o più popolazioni che contengono una certa quantità di variabilità genetica, originatasi tramite le mutazioni, che determinano la comparsa di numerosi alleli ai differenti loci. La variabilità genetica nelle popolazioni è descritta tramite le frequenze alleliche. Le frequenze alleliche ad ogni locus possono variare nel corso delle generazioni a causa delle mutazioni, della selezione naturale, della migrazione o della deriva genetica. Le diverse combinazioni degli alleli che sono presenti ad ogni locus determinano i genotipi individuali, le cui frequenze nelle popolazioni possono essere stimate. In una popolazione ideale, in cui non agiscono forze evolutive, le frequenze genotipiche restano costanti da una generazione all'altra. La genetica delle popolazioni si basa su un modello di popolazione ideale, astratta, che si regge su una serie di assunzioni. Il modello deve essere semplice, in modo da rendere possibile l'analisi matematica, ma, conseguentemente, sarà poco realistico. Ogni modello deve realizzare un compromesso fra semplicità e realismo. In genetica, la popolazione ideale ha dimensioni infinite, gli individui si accoppiano e riproducono "a caso": ogni individuo ha la stessa probabilità di riprodursi con qualsiasi altro individuo della popolazione e la scelta di coppia di un individuo non è influenzata dalla scelta di coppia degli altri individui. In pratica si analizzano campioni di dimensioni limitate (la dimensione dei campioni si indica con n), che sono stati ottenuti tramite campionamenti casuali delle popolazioni reali. Questi campioni sono differenti fra di loro a causa del numero limitato di individui che ogni volta sono effettivamente campionati (campionamento statistico).

La legge di Hardy-Weinberg (1909) definisce la relazione che esiste fra le frequenze alleliche e genotipiche a ciascun locus in una popolazione. La legge di Hardy-Weinberg afferma che, in una popolazione mendeliana, le proporzioni dei genotipi rimangono costanti da una generazione all'altra. In un locus con due alleli (a_1 e a_2), con frequenze p e q (con $p + q = 1$), le frequenze genotipiche sono ricavate dalla proporzione: $a_1a_1:2a_1a_2:a_2a_2 = p^2:2pq:q^2$. Le frequenze di Hardy-Weinberg si possono ricavare sempli-

cemente dall'analisi di una tabella che riporta i risultati di tutte le combinazioni possibili fra i genotipi materni e paterni che sono presenti in una popolazione ad un locus polimorfico con due alleli (Fig. 22).

E' possibile stimare le frequenze genotipiche di una popolazione in

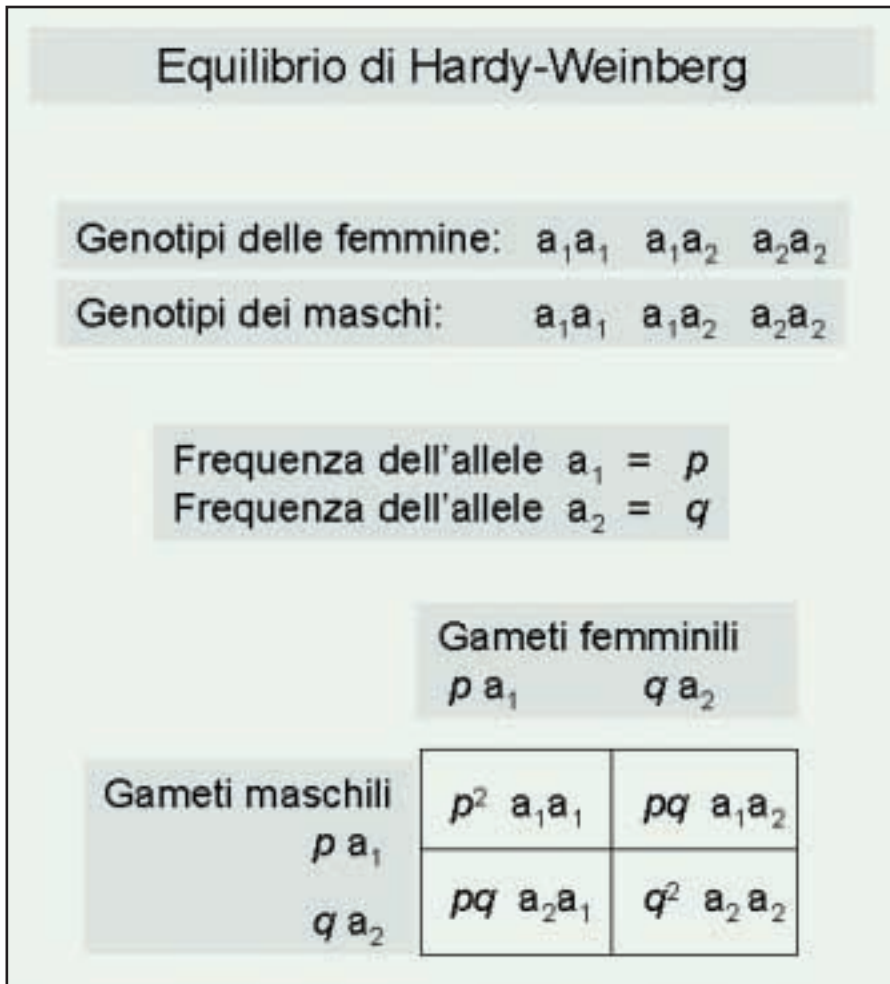


Figura 22 - Equilibrio di Hardy-Weinberg in una popolazione composta da maschi e femmine che hanno due alleli al locus A. Le frequenze degli alleli nella popolazione sono $a_1 = p$, $a_2 = q$. I gameti hanno aplotipi a_1 e a_2 , che sono prodotti con frequenze p e q , rispettivamente. La combinazione casuale dei gameti produce i genotipi della generazione seguente, che avranno frequenze uguali ai prodotti delle frequenze dai gameti.

equilibrio di Hardy-Weinberg (HWE) utilizzando le frequenze alleliche osservate. Per esempio, se sappiamo che l'allele a_1 ha frequenza $p = 0.2$, allora le frequenze genotipiche saranno:

$$\text{genotipo omozigote: } p(a_1a_1) = p^2 = (0.2 \times 0.2) = 0.04$$

$$\text{genotipo eterozigote: } p(a_1a_2) = 2pq = 2 \times (0.2 \times 0.8) = 2 \times 0.16 = 0.32$$

Ovviamente, le frequenze genotipiche attese, vale a dire calcolate tramite le frequenze alleliche, sono rappresentative delle frequenze genotipiche della popolazione se e solo se la popolazione è in HWE. La stessa legge di Hardy-Weinberg stabilisce che le frequenze genotipiche ed alleliche non cambiano dopo la prima generazione di riproduzione casuale. La popolazione è, e resta, in equilibrio di HWE dopo una generazione e per tutte le generazioni seguenti, in assenza di forze evolutive. In certi casi, comunque, una popolazione resta in HWE anche se c'è una certa pressione di selezione naturale, oppure se c'è una certa proporzione di accoppiamenti non casuali.

Nel calcolo delle frequenze genotipiche si applica la legge del prodotto, che implica che ogni evento sia indipendente da tutti gli altri. Se una popolazione non è in HWE (cosa che si può saggiare applicando un test di chi quadrato), la stima delle frequenze genotipiche a partire dalle frequenze alleliche può essere sbagliata. Cause di scostamento da HWE possono essere: accoppiamenti assortativi (non-random mating); migrazione e flusso genico (gene flow); effetto del fondatore (founder effect); colli di bottiglia (bottleneck); deriva genetica (random drift).

In genetica forense di solito si utilizzano sequenze di DNA non codificante, oppure sequenze di DNA ripetitivo. Gli effetti della selezione naturale su queste sequenze (selettivamente neutrali) di solito sono irrilevanti. Inoltre, le analisi di genetica forense hanno lo scopo di stabilire corrispondenze fra un campione ed un individuo da cui il campione origina, oppure di stabilire relazioni di parentela in nuclei familiari. In questi casi gli effetti delle mutazioni sono irrilevanti. Al contrario, le conseguenze dovute alla migrazione ed alle riduzioni delle dimensioni della popolazione possono avere conseguenze importanti, determinando popolazioni miste o popolazioni con alti livelli di inbreeding, che possono essere in disequilibrio di Hardy-Weinberg (HWD).

La migrazione ed il mischiarsi di popolazioni differenziate, origina popolazioni stratificate, che sono geneticamente eterogenee. Se una popolazione è suddivisa in sottogruppi geneticamente distinti, con riproduzione casuale all'interno dei sottogruppi, ma con scarso flusso genico fra i sottogruppi, allora è possibile calcolare le frequenze alleliche e genotipi-

che separatamente nei sottogruppi. Altrimenti le frequenze sono calcolate nella popolazione mista. Se le frequenze alleliche sono diverse nei sottogruppi, allora la popolazione totale può essere in HWD. Questo fenomeno è noto come effetto Wahlund: l'omozigosità osservata in una popolazione composta da due sottogruppi che si sono mischiati recentemente e che non sono panmittici, è significativamente più alta dell'omozigosità attesa sulla base di HWE nella popolazione totale. Sono disponibili procedure di analisi delle popolazioni miste che consentono di identificare il numero di sottopopolazioni presenti e di assegnare ogni individuo alla sottopopolazione di origine (Pritchard *et al.*, 2000).

Inbreeding

Gli individui di una popolazione reale possono riprodursi in modo non casuale perché si scelgono reciprocamente sulla base di certi criteri fenotipici, comportamentali o sociali (accoppiamento assortativo), oppure perché sono fra di loro imparentati. La riproduzione fra individui imparentati si chiama "inbreeding" (consanguineità). Nelle popolazioni reali, tutti gli individui sono in qualche modo imparentati, se si risale per un certo numero di generazioni nel passato. Tuttavia, convenzionalmente, due individui sono considerati inbred se derivano da genitori imparentati nelle tre o quattro generazioni passate. Le conseguenze dell'accoppiamento assortativo e dell'inbreeding sono simili: aumenta la frequenza degli individui omozigoti nella popolazione, e quindi aumenta l'omozigosità rispetto ad una popolazione che sia in HWE. In genetica forense può essere importante conoscere le relazioni di parentela fra gli individui analizzati, perché l'inbreeding può cambiare significativamente le probabilità di identità fra due genotipi.

L'inbreeding negli individui. Due individui che hanno un recente progenitore in comune sono imparentati ed i loro figli sono "inbred" (consanguinei). Le conseguenze genetiche dell'inbreeding derivano direttamente dalle leggi di Mendel. Ogni individuo riceve metà dei suoi alleli da ciascuno dei due genitori e trasmette metà dei suoi alleli a ciascuno dei propri figli. Le probabilità di ricevere oppure di trasmettere l'uno o l'altro dei due alleli presenti ad ogni locus sono equivalenti. Un individuo nato da due genitori imparentati ha una certa probabilità di ricevere entrambi gli alleli ad un locus che sono copie dello stesso allele, cioè sono alleli identici per discendenza (identity by descent: "ibd"). La probabilità che un individuo riceva coppie di alleli ibd dai suoi genitori corrisponde al valore del "coefficiente di inbreeding F ". F è, quindi, una stima della pro-

babilità di omozigosi dovuta a coppie di alleli ibd. E' possibile stimare il valore del coefficiente di inbreeding individuale assumendo che esista una popolazione iniziale di riferimento, in cui tutti gli individui sono non-imparentati. Se tutte le riproduzioni che sono avvenute in questa popolazione sono state registrate, allora esiste un "pedigree" che può essere utilizzato per il calcolo dei valori individuali di F . Per esempio, l'inbreeding di un individuo nato dall'accoppiamento fra due fratelli, che hanno entrambi i genitori in comune, è $F = 0.25$, equivalente alla probabilità che l'individuo riceva una coppia di alleli ibd (Fig. 23). L'inbreeding di un individuo nato dall'accoppiamento fra due fratellastri, che hanno un solo genitore in comune, è $F = 0.125$ (Fig. 24).

Esistono metodi, applicati nella realizzazione di programmi di computer, che consentono di stabilire i valori individuali di F tramite l'analisi di pedigree molto complessi. In caso di pedigree semplici, i valori dei

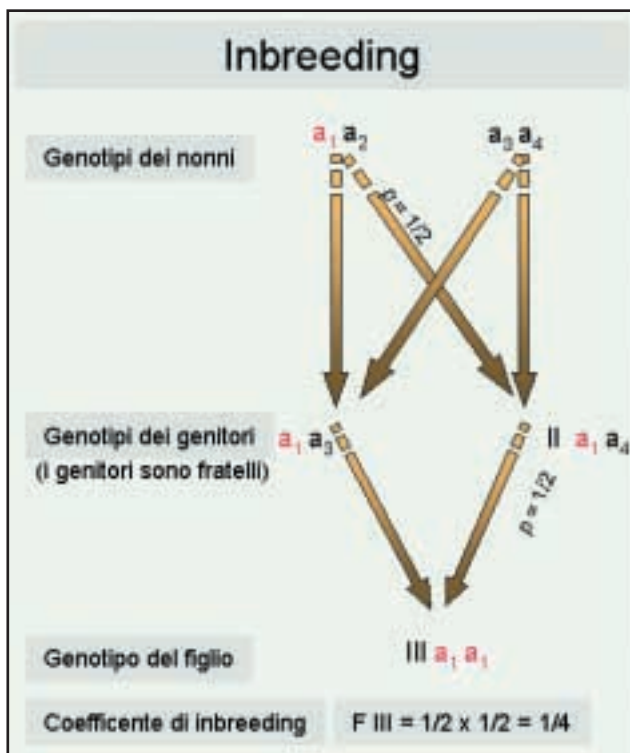


Figura 23 - L'inbreeding di un individuo nato dall'accoppiamento fra due fratelli (entrambi i genitori in comune) è $F = 0.25$ ($1/4$).

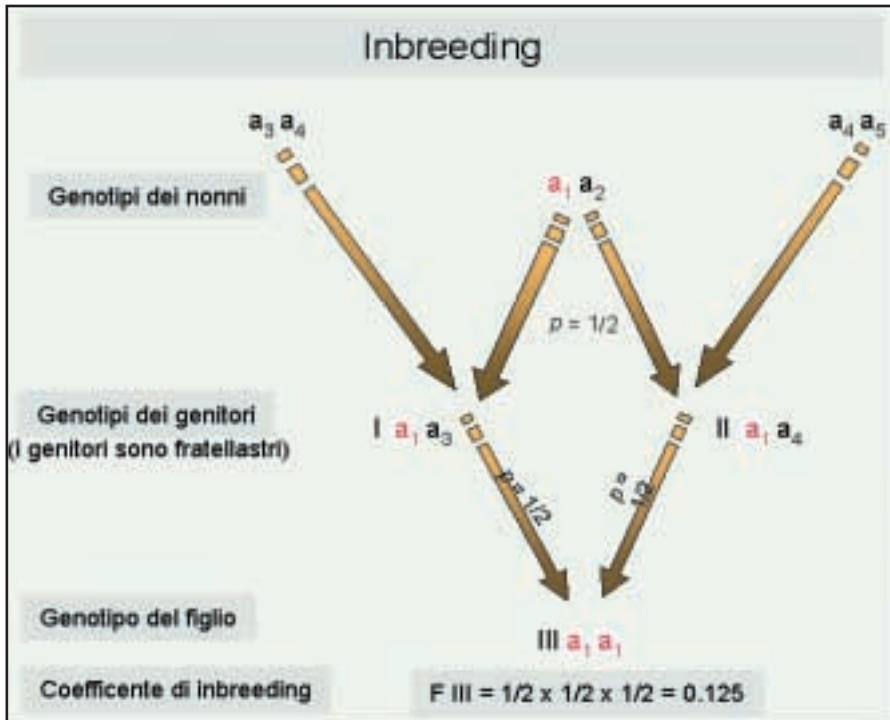


Figura 24 - L'inbreeding di un individuo nato dall'accoppiamento fra due frateLLastri, che hanno un solo genitore in comune, è $F = 0.125$ ($1/8$).

coefficienti di inbreeding fra coppie di individui imparentati si possono direttamente ricavare dalla tabella 1.

Esiste un'altra misura di inbreeding, il coefficiente di coancestry (q), che corrisponde alla probabilità che due alleli ad un locus in due individui presi a caso dalla popolazione siano idb. Il coefficiente di inbreeding F stima la probabilità che un individuo preso a caso nella popolazione riceva ad ogni locus due alleli idb; il coefficiente di coancestry q stima la probabilità che due individui presi a caso abbiano due alleli idb, e quindi è una misura di inbreeding fra individui. Il coefficiente di inbreeding F ed il coefficiente di coancestry q sono correlati. Se due individui, X e Y, hanno un figlio I, $F_I = q_{XY}$, cioè il valore del coefficiente di inbreeding del figlio corrisponde al valore del coefficiente di coancestry fra i suoi due genitori, poiché la probabilità che un locus nel figlio sia omozigote per alleli idb è uguale alla probabilità che i suoi due genitori abbiano due alleli idb al locus. Quindi, $q = 1/4$ per i figli di due fratelli (che hanno $F =$

Tabella 1 - Coefficiente di inbreeding: valori dei coefficienti di inbreeding fra coppie di individui imparentati.

Parentela	Grado	F
Gemelli monozigoti	Identici	
Gemelli dizigoti	Primo	1/4
Fratelli	Primo	1/4
Genitore - figlio	Primo	1/4
Zio - nipote	Secondo	1/8
Fratellastri	Secondo	1/8
Primi cugini	Terzo	1/16
Secondi cugini	Quinto	1/64

0.25), 1/8 per i figli di due fratelli (che hanno $F = 0.125$) e 1/16 per i figli di due primi cugini (che hanno $F = 0.0625$).

Inbreeding nelle popolazioni. In una popolazione il coefficiente medio di inbreeding corrisponde alla probabilità che due alleli ad un locus di un individuo scelto a caso siano ibd. In una popolazione di dimensioni finite c'è una probabilità crescente nel tempo che ogni coppia di alleli sia ibd, semplicemente perché ad ogni

generazione alcuni alleli non saranno trasmessi alla generazione seguente, mentre altri saranno trasmessi in copie multiple. Se una popolazione ha inbreeding F , quale sarà il valore di F nella generazione seguente? Assumiamo di avere una popolazione composta da N individui. La probabilità che un individuo della generazione seguente abbia due genitori distinti è: $1 - 1/N$, e la probabilità che riceva due alleli ibd corrisponde al valore di q . Si può dimostrare che, dopo una generazione, il coefficiente di inbreeding della popolazione è:

$$F' = q' = 1/2N + (1 - 1/2N)q$$

Alla generazione t :

$$qt = 1 - (1 - 1/2N)^t$$

In figura 25 si mostrano i valori di $F = q$ nel corso di t generazioni per popolazioni di $N = 100.000$, 10.000 e 1.000 individui (da Evett and Weir, 1998). L'incremento dell'inbreeding accompagna una diminuzione di variabilità genetica nella popolazione, a causa del drift. Chiaramente, sia l'aumento di F sia la diminuzione di variabilità genetica sono piccoli in caso di N grande. I valori di inbreeding raggiungeranno il valore di 1 dopo circa 10.000 generazioni in una popolazione di $N = 1.000$ individui, e dopo circa 10.000.000 in una popolazione di $N = 100.000$.

Occorre notare che N è la "dimensione effettiva" della popolazione, corrispondente al numero degli individui che effettivamente si riproducono e che possono così trasmettere i loro geni alle generazioni seguenti. Quasi sempre la dimensione effettiva è più piccola o molto più piccola

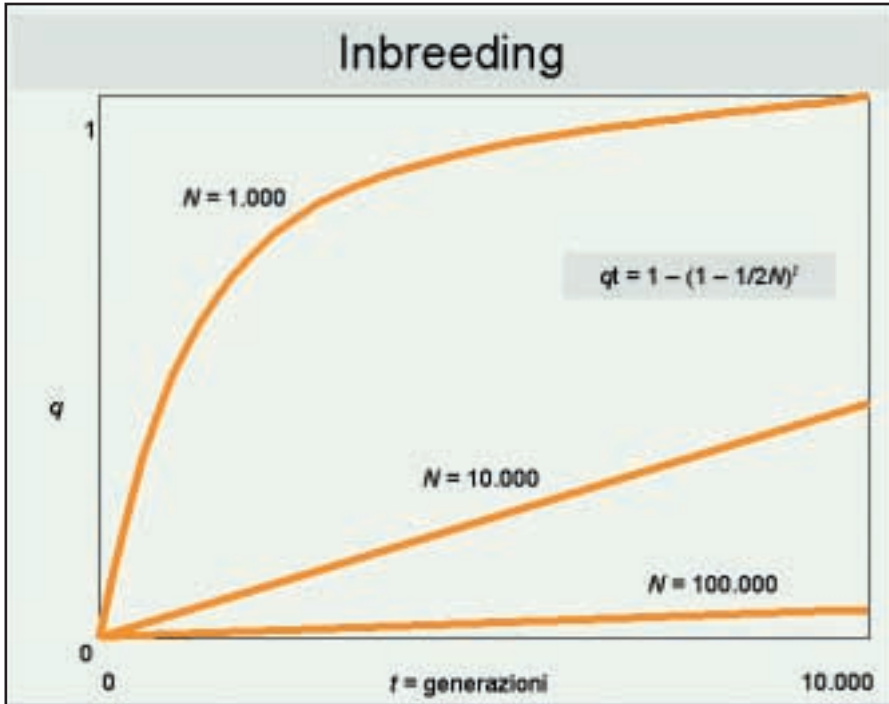


Figura 25 - Variazione dei valori di F (coefficiente di inbreeding) = q (coefficiente di coancestry) nel corso di t generazioni per popolazioni di $N = 100.000$, 10.000 e 1.000 individui (da Evett and Weir, 1998). I valori di inbreeding raggiungono il valore di 1 (inbreeding totale) dopo circa 10.000 generazioni in una popolazione di $N = 1.000$ individui, e dopo circa 10.000.000 in una popolazione di $N = 100.000$.

della dimensione censita, cioè del numero di individui che sono presenti nella popolazione. Questo avviene perché non tutti gli individui presenti in una popolazione rientrano contemporaneamente nelle classi di età riproduttive, perché il rapporto fra i sessi nelle classi di età riproduttive non è sempre corrispondente ad 1 maschio per 1 femmina, perché la fertilità e fecondità degli individui non è identica, perché la sopravvivenza delle famiglie di fratelli non è identica, e così via. La dimensione effettiva può essere da 2 a 20 volte inferiore alla dimensione osservata nella popolazione. Le popolazioni isolate (che non ricevono immigranti) di piccole dimensioni effettive sono soggette all'azione della deriva genetica ("drift"): le frequenze alleliche fluttuano da una generazione all'altra in conseguenza del campionamento di un piccolo numero di gameti sul totale dei gameti potenzialmente prodotti dalla generazione parentale. A

lungo termine il drift produce una perdita di alleli e quindi una diminuzione della variabilità genetica nella popolazione. Infatti, le fluttuazioni casuali delle frequenze alleliche portano alla perdita di alleli, con conseguente fissazione degli alleli alternativi. I loci fissati diventano monomorfici e il valore di eterozigosità si azzerava.

Il drift produce inevitabilmente un aumento di F nelle popolazioni reali, che hanno sempre dimensioni finite, ed hanno spesso dimensioni effettive molto più piccole delle dimensioni osservate. Per rendere più realistico il modello di genetica delle popolazioni, può essere opportuno esprimere le frequenze genotipiche attese sulla base di HWE, in termini sia delle frequenze alleliche sia del coefficiente di inbreeding. L'inbreeding provoca un aumento della frequenza dei genotipi omozigoti ed una diminuzione della frequenza dei genotipi eterozigoti, cioè una diminuzione dell'eterozigosi H . Un individuo omozigote riceve dai genitori due copie dello stesso allele a . In una popolazione finita esiste una probabilità F che queste copie dell'allele a siano **ibd**. La probabilità che le copie di a siano **non-ibd** sarà naturalmente $(1 - F)$. Perciò la frequenza di questo genotipo può essere espressa come frequenza attesa di HWE, pesata per una componente dovuta al drift (o dovuta a qualsiasi altra causa che aumenta l'inbreeding). Il coefficiente di inbreeding F misura la diminuzione di H rispetto all'eterozigosità attesa in una popolazione in HWE, che corrisponde a $2pq$, cioè:

$$F = (2pq - H) / 2pq$$

Da cui deriva:

$$H = 2pq(1 - F).$$

In una popolazione inbred le frequenze attese dei tre genotipi ad un locus con due alleli sono:

$$p(a_1a_1) = Fp + (1 - F)p^2$$

$$p(a_1a_2) = 2pq(1 - F)$$

$$p(a_2a_2) = Fq + (1 - F)q^2$$

Per esempio: se l'allele a_1 ha frequenza $p = 0.05$, la frequenza del genotipo omozigote in una popolazione in HWE sarà $p^2 = 0.0025$. Ma in un gruppo di individui che derivano da accoppiamenti fra cugini, con $F = 1/16$ e $(1 - F) = 15/16$, la frequenza dell'omozigote sarà:

$$1/16 \times 0.05 + 15/16 \times 0.0025 = 0.003 + 0.0023 = 0.0053$$

cioè, sarà più che raddoppiata rispetto ad una popolazione in HWE.

Se la popolazione è in HWE e $F = 0$ (non c'è inbreeding), le tre equazioni

precedenti diventano uguali alle frequenze alleliche attese per i tre genotipi: p^2 , $2pq$, q^2 . Al contrario, se $F = 1$, la popolazione risulta composta esclusivamente dai due genotipi omozigoti con frequenze p e q , rispettivamente.

Da questo modello di stima dell'inbreeding è possibile sviluppare derivazioni più complesse. Possiamo considerare le relazioni fra i genotipi di coppie di individui imparentati, per esempio fratelli, fra individui che appartengono a sottopopolazioni distinte, ecc. E' possibile stimare le probabilità che due individui non inbred abbiano 0, 1, 2, ecc. paia di alleli ibd, la probabilità che due individui abbiano due loci entrambi omozigoti, o che abbiano un locus omozigote ed uno eterozigote, oppure abbiano gli stessi o differenti genotipi eterozigoti ai due loci. Inoltre, è possibile stimare le probabilità che due individui imparentati derivati da genitori non-inbred e non imparentati, abbiano lo stesso genotipo.

GENETICA MOLECOLARE: METODI DI ANALISI DELLA VARIABILITÀ GENETICA A LIVELLO DEL DNA

Le procedure d'analisi molecolare utilizzate in genetica delle popolazioni ed in genetica forense, consistono in: raccolta e conservazione dei campioni biologici; estrazione del DNA; digestione del DNA tramite enzimi di restrizione; separazione dei frammenti del DNA tramite elettroforesi in gel di agarosio o di acrilamide; immobilizzazione dei frammenti di DNA tramite Southern blotting; preparazione di sonde nucleotidiche marcate; ibridazione ed identificazione dei frammenti di restrizione; amplificazione del DNA tramite PCR; sequenziamento nucleotidico; analisi automatizzata delle sequenze, dei microsatelliti e dei frammenti di DNA (Fig. 26).

Metodi di raccolta dei campioni biologici

Le tecniche d'analisi molecolare che si basano sulla PCR richiedono solo piccole quantità di DNA e, quindi, consentono l'utilizzo di qualsiasi tipo di campione biologico. Tuttavia, queste tecniche sono molto esposte ai rischi di contaminazione. Le tecniche più tradizionali (ad esempio, l'analisi dei frammenti di restrizione) richiedono invece quantità maggiori di DNA che si possono ottenere solo dalla disponibilità di campioni biologici che non siano troppo piccoli o troppo degradati. In ogni caso, le procedure d'analisi e la qualità dei risultati sono dipendenti dalla qualità dei campioni e dalle eventuali contaminazioni. E' quindi necessario rac-

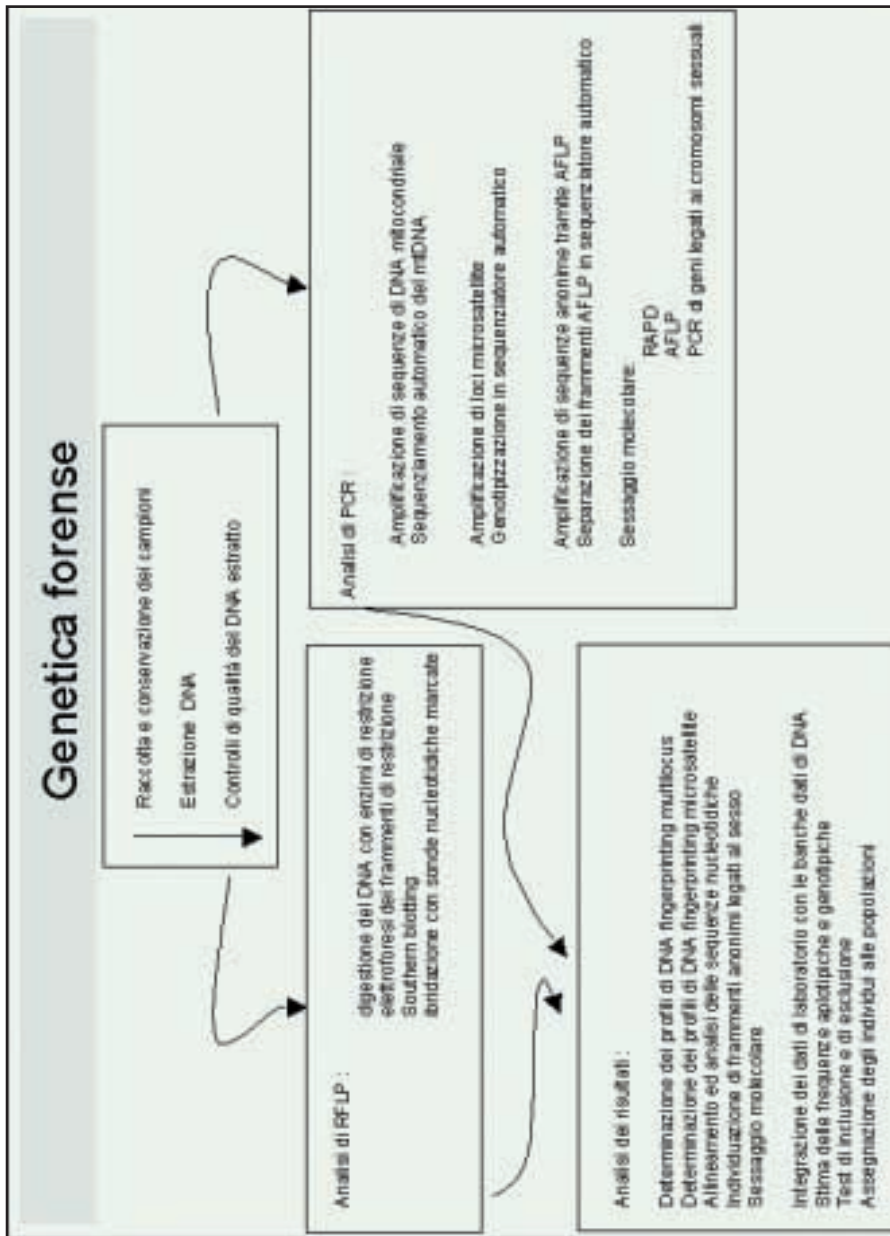


Figura 26 - Le procedure d'analisi molecolare utilizzate in genetica delle popolazioni ed in genetica forense, consistono nella raccolta e conservazione dei campioni biologici, estrazione del DNA, digestione del DNA tramite enzimi di restrizione, separazione dei frammenti del DNA tramite elettroforesi in gel di agarosio o di acrilamide, immobilizzazione dei frammenti di DNA tramite Southern blotting, preparazione di sonde nucleotidiche marcate, ibridazione ed identificazione dei frammenti di restrizione, amplificazione del DNA tramite PCR, sequenziamento nucleotidico, analisi automatizzata delle sequenze, dei microsatelliti e dei frammenti di DNA.

cogliere e conservare i campioni biologici nel modo migliore possibile.

I campioni d'origine animale più utilizzati sono i seguenti:

Campioni di sangue prelevati da animali vivi. Il DNA nucleare è ricavato dalle cellule bianche del sangue, che sono dotate di nucleo, oppure dai globuli rossi, che sono nucleati negli anfibi, rettili ed uccelli. I mammiferi hanno globuli rossi anucleati. La maggior parte del DNA mitocondriale (mtDNA) è ricavato dalle cellule bianche. Il DNA necessario ad effettuare le diagnosi molecolari che sono usate in genetica forense si può ottenere da prelievi di circa 10 - 100 microlitri (μl) di sangue di anfibi, rettili ed uccelli, oppure da circa 0.5 - 1.0 millilitri (ml) di sangue di mammiferi. Il mtDNA è presente in copie multiple nel citoplasma delle cellule bianche e può essere ricavato da pochi μl di sangue, indipendentemente dalla specie. Dai prelievi di sangue normalmente si estrae il DNA totale (che contiene sia mtDNA che DNA nucleare). I geni nucleari o mitocondriali sono poi amplificati selettivamente tramite PCR, oppure individuati selettivamente tramite l'uso di sonde nucleotidiche specifiche. Il sangue può essere prelevato semplicemente pungendo una vena e raccogliendo le gocce con un capillare, oppure tramite siringhe sterili tipo insulina. Le siringhe tipo insulina consentono di prelevare da pochi μl fino ad 1 ml di sangue utilizzando un ago di diametro molto sottile. Il prelievo può essere effettuato da qualsiasi vaso venoso che sia ben accessibile. Il prelievo deve essere effettuato con cura, in modo da evitare conseguenze dannose per l'animale. In ogni caso il prelievo deve essere preceduto dalla disinfezione dell'epidermide, e deve essere effettuato utilizzando ogni volta una siringa nuova, sterile e non riciclata. I prelievi per le analisi CITES devono essere effettuati dal veterinario di fiducia del proprietario degli animali. I costi delle operazioni di prelievo sono a carico dei proprietari. I prelievi possono essere effettuati utilizzando soluzioni anti-coagulanti. Si consiglia di usare il sale bisodico dell'EDTA (EDTA.Na_2), sostanza anticoagulante che non interferisce con le procedure di estrazione ed analisi del DNA. Poche gocce di una soluzione al 10% di EDTA.Na_2 (ottenuta sciogliendo 10 grammi di EDTA.Na_2 in 100 ml di acqua bidistillata sterile tiepida) sono sufficienti per un prelievo di sangue di 1 ml effettuato utilizzando una siringa da insulina. I campioni di sangue sono conservati in tampone di Longmire (LongBuffer).

Campioni di tessuti solidi prelevati da animali vivi (tramite biopsie) o morti. Piccole biopsie, effettuate ad esempio dalle orecchie, oppure campioni di circa 0.5 - 2.0 gr di tessuto prelevati dalle carcasse degli animali morti, sono sufficienti per effettuare le analisi del DNA. I campioni di tessuto, di qualsiasi tipo, devono essere conservati in provette di plastica

sterili a chiusura ermetica contenenti etanolo (alcool etilico) al 90 - 100% (EtOH 100%). E' estremamente importante utilizzare etanolo puro, cioè etanolo per laboratorio non denaturato. Le sostanze denaturanti, che colorano in rosa l'alcool, possono contaminare il DNA e rendere impossibili le analisi genetiche. E' pure estremamente importante conservare i campioni in volumi di etanolo almeno 10 volte superiori al peso del tessuto (per esempio, 1 gr di tessuto deve essere conservato in almeno 10 ml di etanolo). L'etanolo disidrata i tessuti, ed in questo modo blocca le reazioni biochimiche che porterebbero alla degradazione del DNA. I tessuti contengono acqua che, nel corso della disidratazione, diluisce l'etanolo. Perciò, per evitare un'eccessiva diluizione, è necessario usare volumi abbondanti di etanolo al 90 - 100%. Il DNA è stabile in etanolo a temperatura ambiente. I campioni in etanolo, pertanto, possono essere conservati a temperatura ambiente, oppure refrigerati a qualsiasi temperatura inferiore alla temperatura ambiente. Gli animali morti destinati alle analisi del DNA devono essere congelati immediatamente ed i corpi devono essere conservati alla più bassa temperatura possibile. Normalmente, il congelamento a -10 / -15°C garantisce la conservazione del DNA per alcuni mesi, mentre il congelamento a -80°C, oppure in azoto liquido, consente la conservazione del DNA per molti anni. In ogni caso, occorre tener presente che il congelamento di corpi di animali di grosse dimensioni avviene lentamente, a partire dalle superfici esterne verso l'interno. Il DNA dei tessuti interni può andare incontro a processi di degradazione se il processo di congelamento si prolunga nel tempo. Occorre aver ben presente che i congelatori sono soggetti a rompersi, e che l'erogazione di energia elettrica può interrompersi. Prolungati o ripetuti scongelamenti dei tessuti producono la degradazione del DNA. Si suggerisce, pertanto, di effettuare quanto prima possibile i prelievi delle aliquote di tessuto destinati alle analisi genetiche e di non conservare i corpi congelati a lungo. Le aliquote di tessuto destinate alle analisi genetiche devono essere conservate in etanolo. Le biopsie ed i prelievi devono essere effettuati facendo estrema attenzione alle contaminazioni. Occorre lavorare su superfici pulite, disinfettando o lavando le parti dal corpo da cui si effettuano i prelievi, utilizzando bisturi, forbici e pinze pulite (del tipo usa-e-getta), e sterilizzate alla fiamma, evitando di toccare i campioni con le dita ed eventualmente utilizzando guanti di lattice sterili da laboratorio.

Campioni di penne e peli. Si possono ottenere quantità di DNA sufficienti ad effettuare analisi tramite PCR, utilizzando le cellule che sono presenti nei bulbi piliferi e nei calami (radici) delle penne. In questo caso occorre prelevare circa 10 - 20 peli con i bulbi, oppure 2 - 4 piume o

penne da ogni esemplare. Il prelievo si può fare utilizzando guanti da laboratorio, oppure con una pinza, curando attentamente di evitare di toccare o sporcare in qualsiasi modo le radici dei peli o delle penne. Si possono utilizzare peli e penne prelevate da animali morti e congelati, purché le condizioni di congelamento siano buone, come si è specificato sopra. Si possono utilizzare campioni di peli o penne raccolte sul terreno o nelle gabbie che ospitano gli animali. I campioni di penne e peli sono conservati in EtOH 100%.

Campioni di ossa, scaglie, squame. Si possono utilizzare campioni di circa 2 - 4 gr di tessuto osseo o di scaglie e squame, che sono prelevati e trattati come descritto per i campioni di tessuti, peli o penne. Questi campioni possono essere conservati in congelatore o in EtOH 100%.

Campioni non-invasivi. La PCR consente di amplificare selettivamente il DNA estratto da campioni di escrementi, saliva, ecc. Questi campioni sono esposti alla rapida degradazione e contaminazione del DNA, e devono essere raccolti con molta cura.

Metodi di raccolta delle tracce

Le tracce biologiche (campioni di sangue od altri fluidi biologici depositi su substrati solidi, tipo stoffe, foglie, sassi, cortecce, ecc.), o altri campioni raccolti nel corso delle indagini ed accertamenti, spesso sono essiccati e non freschi. Questi campioni contengono poco DNA, che può essere degradato o contaminato da DNA esogeno. Le tracce possono essere raccolte direttamente o rimosse tramite l'uso di un supporto appropriato. Una traccia può essere raccolta direttamente usando, per esempio, pinze sterili e guanti da laboratorio, e deposta in un contenitore appropriato (sacchetto di plastica sterile e ben chiudibile; provetta da laboratorio sterile). Il DNA contenuto nelle tracce è sempre esposto alla degradazione e quindi ai rischi di contaminazione. Perciò, nel raccogliere questi campioni devono essere usate tutte le precauzioni possibili per evitare contaminazioni, particolarmente con la pelle, capelli, saliva, ecc., di chi effettua il campionamento. Le tracce vecchie possono essere conservate in sacchetti di plastica o contenitori sterili, congelate a - 20°C. Le tracce fresche probabilmente contengono DNA non degradato e quindi devono essere disposte in provette o altri contenitori con EtOH 100%. Le tracce in etanolo possono essere conservate a temperatura ambiente, o refrigerate a qualsiasi temperatura inferiore alla temperatura ambiente. Il DNA contenuto nelle tracce e nei campioni biologici è stabile per anni in EtOH 100% conservato a temperatura ambiente o refrigerato. Se la

raccolta diretta non è possibile, una traccia può essere rimossa e trasferita su un substrato solido, ad esempio su carta assorbente. La traccia può essere raschiata utilizzando la lama sterile di un bisturi tipo usa-e-getta, o può essere reidratata utilizzando alcune gocce di acqua sterile oppure di soluzione fisiologica, e poi assorbita in carta assorbente o cotone idrofilo. Una traccia reidratata in acqua è molto più esposta alla degradazione del DNA di una traccia secca. Una traccia reidratata in acqua deve essere al più presto trasferita in congelatore, oppure in tampone tipo LongBuffer. Il DNA contenuto in una traccia reidratata in una soluzione tampone appropriata è stabile. Tamponi tipo LongBuffer preservano intatto il DNA a temperatura ambiente per un certo periodo (alcune settimane) o congelato (per sempre). Quando si procede al recupero delle tracce, è buona norma raccogliere anche uno più campioni dalle aree adiacenti che non mostrano apparenti tracce biologiche. Questi campioni servono come controllo per determinare cosa fosse presente sul substrato prima che la traccia fosse deposta, e possono facilitare l'identificazione di eventuali DNA contaminanti.

Conservazione dei campioni

Ogni campione biologico è soggetto a degradazione se non è raccolto e conservato appropriatamente. Sebbene il DNA sia molto più stabile delle proteine e degli enzimi, è comunque soggetto a degradazione, sia a causa dell'azione digestiva delle endonucleasi, enzimi litici che sono presenti normalmente nelle cellule e che sono attivati nel corso della morte cellulare, sia a causa dell'azione di agenti esterni (agenti biologici: muffe, funghi, batteri; agenti fisici: luce solare ed ultravioletta, temperatura, umidità). La quantità e qualità del DNA dipende essenzialmente dalle condizioni di conservazione dei campioni biologici da cui il DNA è estratto. In conseguenza della degradazione i lunghissimi filamenti integri del DNA dei cromosomi (che possono essere lunghi fino ad alcuni metri) vengono frammentati in segmenti di poche decine o centinaia di nucleotidi. Il DNA degradato e frammentato non è più analizzabile utilizzando metodi come l'RFLP, mentre, entro certi limiti, resta analizzabile tramite PCR. Deve essere sottolineato come la degradazione comunque non cambia le caratteristiche delle sequenze del DNA. La degradazione limita le possibilità di analizzare il DNA, ma non ne invalida i risultati. L'analisi del DNA degradato è maggiormente esposta alle contaminazioni con DNA esogeno ed alla produzione di artefatti. Ad esempio, è possibile che la PCR di un campione di DNA degradato produca l'amplificazione di uno

solo dei due alleli che sono presenti ad un locus eterozigote (“dropout allelico”). E’ possibile che l’allele di maggior peso molecolare, determinato da una sequenza di DNA più lunga, sia degradato, cioè frammentato, e non possa venire amplificato. In questo caso un locus eterozigote viene confuso per omozigote. Esistono procedure di controllo di qualità che consentono di individuare e correggere il dropout allelico.

Conservazione dei campioni vegetali. Alcune specie vegetali producono composti quali tannini, fenoli ed altri metaboliti secondari, che possono interferire con l’estrazione del DNA. E’ possibile estrarre DNA da piante conservate negli erbari, ma è indubbiamente consigliabile usare il materiale più fresco possibile. I campioni vegetali, costituiti di solito da foglie oppure germogli, appena raccolti devono essere conservati freschi ed umidi, ad esempio deposti su uno strato di ghiaccio, oppure possono essere immediatamente congelati a - 20°C, a - 80°C, o in azoto liquido. I campioni congelati non devono essere scongelati prima dell’inizio dell’estrazione del DNA. Qualora non fosse possibile congelare, allora occorre desiccare i campioni rapidamente in contenitori individuali che contengono gel di silica (tipo Sigma S7500 o S7625). Per conservare campioni di circa 1 gr di tessuto fogliare si possono usare tubi di plastica da 15 ml contenenti circa 5 gr di silica posti sotto un batuffolo di cotone. Occorre controllare che il silica non sia saturato dall’umidità del campione. La saturazione è segnalata dal cambio di colore dal blu al rosa dei granuli di silica. I tessuti vegetali conservati in silica possono essere mantenuti a temperatura ambiente per un periodo indefinito. Probabilmente la desiccazione è il migliore fra i metodi di conservazione dei campioni vegetali. Esiste, inoltre, almeno un metodo chimico che consente buone estrazioni di DNA da campioni di foglie di specie come le querce che contengono tannini e fenoli. Questi composti non sono degradati dal congelamento e possono venire co-purificati col DNA, interferendo con le successive analisi molecolari. Questo metodo utilizza una soluzione di NaCl-CTAB: si prepara una soluzione d’acqua distillata sterile saturata con cloruro di sodio (NaCl). Si aggiunge NaCl finché si forma un fondo insolubile di circa 1 cm. Poi si aggiunge il CTAB (bromuro di esadeciltrimetilammonio), lentamente, mischiando finché la soluzione acquista una viscosità simile a quella dell’olio da motore (quest’operazione richiede qualche ora). Occorrono circa 30 - 40 gr per un litro di soluzione satura di NaCl. La quantità esatta di CTAB non è importante. Le foglie sono tagliate in frammenti di circa 1 cm² e immerse immediatamente nella soluzione, rispettando il rapporto di circa una parte di foglie per tre parti di soluzione. I campioni vegetali in CTAB possono essere conservati a tempera-

tura ambiente per almeno un mese e congelati a -20°C indefinitamente.

Metodi di conservazione dei campioni animali. I tessuti di origine animale possono essere conservati congelandoli immediatamente a -20°C , a -80°C , o in azoto liquido. Alternativamente, i tessuti possono essere conservati indefinitamente a temperatura ambiente, oppure refrigerati a qualsiasi temperatura inferiore alla temperatura ambiente, in EtOH 100%, o in tampone DMSO 20%. Questo tampone si prepara sciogliendo circa 200 gr di DMSO (dimetilsulfoxide) in un litro di acqua saturata con NaCl (vedi preparazione del tampone CTAB). Il rapporto fra volume di tessuto e volume di EtOH 100% o di DMSO 20% è di 1:10. I campioni di sangue sono prelevati da animali vivi utilizzando una sostanza anticoagulante (ad esempio, EDTA.Na₂ 10%). Il sangue intero può essere conservato a temperatura ambiente, oppure in frigorifero in tampone di Longmire (LongBuffer). Per preparare 1 litro di LongBuffer, si sciolgono: 37,2 gr di EDTA.Na₂, 0,58 gr di NaCl, 5 gr di SDS (sodiodecilsolfato), in 100 ml di 1M Tris/HCl, a pH 8.0. Quando tutte le sostanze sono in soluzione si aggiunge acqua distillata sterile fino al volume totale di 1 litro. La conservazione del DNA si basa sugli effetti chelanti dell'EDTA e sull'azione proteolitica dello SDS. Il rapporto fra volume di sangue e volume di LongBuffer deve essere di 1:5. I tamponi CTAB, DMSO e LongBuffer sono stabili e si possono conservare a temperatura ambiente per almeno un anno. Il CTAB e DMSO sono tossici ed irritanti e devono essere manipolati con prudenza. L'EtOH 100% ed il LongBuffer non sono tossici. Il Laboratorio di Genetica dell'Istituto Nazionale per la Fauna Selvatica (INFS, Via Cà Fornacetta n. 9, 40064 Ozzano dell'Emilia, Bologna; telefono: 051 6512111; fax: 051 796628; e-mail: met0217@iperbole.bo.it) fornisce gratuitamente provette contenenti EtOH 100% e LongBuffer pronte per la conservazione di campioni di tessuti e di sangue, a chiunque ne faccia motivata richiesta.

Contaminazioni. Ci sono diversi tipi di contaminazione che possono interferire con i risultati delle analisi. Le contaminazioni d'origine non biologica sono dovute alla presenza di sostanze contaminanti (coloranti, saponi ed altre molecole chimiche) che possono inibire l'azione degli enzimi di restrizione o della Taq polimerasi, e quindi impedire le analisi molecolari. Le contaminazioni di origine biologica sono dovute alla presenza di micro-organismi o di sostanze biologiche di altra origine che possono mischiarsi ai campioni nelle varie fasi del procedimento di analisi (nel momento della raccolta del campione, dell'estrazione del DNA, o in alcune fasi delle analisi di laboratorio). Le tecniche d'analisi che usano la PCR sono particolarmente sensibili alle contaminazioni di laboratorio.

L'organizzazione del laboratorio e l'attenzione degli analisti possono evitare le contaminazioni di laboratorio.

Estrazione del DNA

I campioni biologici in etanolo, in soluzione tampone o essiccati, sono sempre conservati in congelatore dal momento in cui sono consegnati al laboratorio di genetica forense. Aliquote dei campioni sono subito utilizzate per l'estrazione del DNA. Il DNA in soluzione sterile di tampone Tris/EDTA (TE) è molto stabile a temperatura ambiente, oppure refrigerato a qualsiasi temperatura inferiore alla temperatura ambiente.

Il DNA può essere estratto da qualsiasi tipo di campione biologico vegetale od animale. In genetica forense i principali problemi derivano dalla degradazione del DNA e dalla contaminazione con DNA esogeno dei campioni da analizzare. I metodi d'estrazione devono consentire di ottenere soluzioni di DNA privo di contaminanti ed impedire ulteriori degradazioni durante le procedure di laboratorio. L'approccio più utilizzato consiste nell'estrarre il DNA totale contenuto in un campione, che include il DNA nucleare ed il DNA mitocondriale, più l'eventuale DNA esogeno dovuto alla presenza di virus e batteri, ed il DNA contaminante di varia origine. Nel corso delle successive procedure di laboratorio si evidenziano, oppure si selezionano le sequenze da analizzare tramite l'uso di sonde specifiche o tramite PCR. Esistono metodi che consentono di separare i mitocondri dai nuclei cellulari e di purificare il DNA mitocondriale. E' anche possibile isolare i singoli cromosomi ed analizzare i geni che mappano in particolari regioni cromosomiche. Tuttavia questi metodi richiedono DNA di buona qualità o l'uso di cellule integre, cose che difficilmente sono disponibili in genetica forense.

Esistono innumerevoli procedure che sono utilizzabili per l'estrazione del DNA totale. Le procedure d'estrazione del DNA prevedono un primo trattamento di lisi delle membrane cellulari e di proteolisi. Questi trattamenti hanno lo scopo di disintegrare tutte le strutture proteiche cellulari e di liberare in soluzione il DNA. Poi viene effettuata una seconda serie di trattamenti per separare il DNA da tutti i residui della degradazione delle strutture proteiche, allo scopo di ottenere una soluzione di DNA che sia il più possibile priva da ogni altra sostanza biologica. A questo punto il DNA può essere recuperato e risospeso in soluzione tampone a concentrazione nota. Le soluzioni tampone usate per l'estrazione per la conservazione del DNA sono a base di Tris, ed hanno la funzione di mantenere un valore di pH costante che inibisce l'azione degli enzimi che degradano il DNA. Per

esempio, le DNasi, gli enzimi cellulari che degradano il DNA, hanno pH ottimali attorno a 7.0, quindi i tamponi di estrazione vengono preparati in modo da mantenere valori di pH fra 8.0 e 9.0. Questi tamponi contengono EDTA che agisce come chelante degli ioni bivalenti di calcio e magnesio, e quindi contribuisce ad inibire ulteriormente l'attività delle DNasi (che richiedono la presenza di questi ioni). I tamponi di digestione contengono Proteinasi K, un enzima che produce la digestione enzimatica delle strutture proteiche, oppure GUS (tiocianato di guanidina), che produce la disintegrazione chimica delle strutture proteiche. L'azione della proteinasi K o del GUS è coadiuvata dalla presenza di SDS, un detergente anionico che solubilizza le membrane cellulari e denatura le proteine.

In genetica forense è molto importante utilizzare metodi che garantiscono l'estrazione ed il recupero della maggior parte del DNA presente nei campioni biologici. La scelta del metodo ottimale di estrazione può essere anche condizionata dall'uso che s'intende fare del DNA. Per esempio, se il DNA deve essere amplificato tramite PCR, allora l'estrazione può essere orientata all'utilizzo di piccoli campioni per ottenere quelle piccole quantità di DNA che sono sufficienti per le reazioni di amplificazione. Altre tecniche, tipo le analisi di RFLP tramite digestione con enzimi di restrizione e Southern blotting, richiedono quantità più grandi di DNA di buona qualità, cioè integro. Una comparazione schematica di alcuni fra i metodi di estrazione più usati, è riportata nella tabella 2. I dettagli delle tecniche possono essere ricavati dalla lettura dei molti manuali disponibili.

La semplice bollitura del campione non elimina i residui cellulari e molecolari che possono inibire la PCR. Il DNA liberato in soluzione è una piccola frazione del DNA totale contenuto nel campione ed è certamente poco pulito. Questa tecnica deve essere utilizzata solo per campioni di tessuto in buone condizioni di conservazione, che devono essere analizzati molto rapidamente utilizzando metodi d'analisi particolarmente robusti,

Tabella 2 - Metodi di estrazione del DNA.

Protocollo	Tessuto	Qualità del DNA
Bollitura	Cellule; tessuti morbidi	Denaturato e contaminato
Chelex	Qualunque; peli; penne	Denaturato
CTAB	Vegetali	Buona
Proteinasi/fenolo/cloroformio	Qualunque	Buona
GUS/silica	Qualunque; forensi, non-invasivi	Buona

come ad esempio, l'amplificazione di una sequenza che amplifica bene in qualsiasi circostanza e che non è sensibile alla presenza di contaminanti. Il Chelex è un metodo rapido ed economico che consente di estrarre DNA di buona qualità, ma che è esposto ad una frammentazione abbastanza rapida che viene prodotta dall'azione delle resine chelanti. Pertanto, l'estrazione col Chelex dovrebbe essere fatta solo per campioni di DNA che devono essere utilizzati immediatamente e che non sono destinati ad essere conservati a lungo. Il Chelex può essere particolarmente utile per preparare campioni di DNA estratti da tracce di sangue o da singoli bulbi piliferi, cioè da campioni che contengono comunque poco DNA che sarà completamente utilizzato per immediate analisi tramite PCR. Se è necessario archiviare i campioni in banche del DNA, oppure se è ipotizzabile che le analisi possano continuare per un certo periodo, allora è consigliabile ricorrere ad altri metodi di estrazione. Le estrazioni in tampone CTAB sono particolarmente utili per eliminare metaboliti secondari di origine vegetale. Questi protocolli sono pertanto utilizzati per estrarre DNA dai campioni freschi di piante e possono essere anche utilizzati per campioni animali che, come gli escrementi, possono contenere abbondanti residui vegetali derivati dall'alimentazione. Il metodo classico della digestione dei tessuti con Proteinasi K, cui seguono ripetute purificazioni del DNA in soluzione tramite estrazione con fenolo e cloroformio, produce DNA di ottima qualità in quasi tutti i casi. Tuttavia, il fenolo ed il cloroformio sono sostanze tossiche, che vanno manipolate con prudenza. Il metodo di digestione dei tessuti con GUS ed estrazione e purificazione del DNA con microgranuli di silice, funziona ottimamente con ogni tipo di tessuto, consente di estrarre e recuperare quasi tutto il DNA presente e non utilizza sostanze tossiche. I metodi che usano il GUS sono ottimi sostituti del metodo classico. Questi metodi producono DNA di ottima qualità e quantità, e sono i metodi di scelta per l'estrazione da campioni problematici. Questi protocolli sono applicati in molti kit commerciali, alcuni dei quali sono forniti in varianti specializzate per l'estrazione da tessuti, sangue, escrementi o campioni museali.

Nel Laboratorio di Genetica dell'INFS si utilizzano i seguenti protocolli di estrazione del DNA.

(1) Estrazione con Chelex da radici di peli e penne

Preparazione delle soluzioni:

- Chelex 5%: si sospendono 2,5 gr di Chelex in 50 ml di acqua bi-distillata sterile (ddH₂O);
- Proteinasi K: si sciolgono 10 mg di Proteinasi K per ogni ml di ddH₂O;

Preparazione dei campioni:

1. Si identificano le provette, del tipo “eppendorf” da 1,5 ml, con il numero del campione;
2. Si trasferiscono 300 μ l di Chelex 5% + 20 ml di Proteinasi K in ciascuna provetta;
3. Si lava il calamo di 1 penna (circa 1 cm), oppure 1 - 10 peli con i bulbi, con ddH₂O; si taglia la penna con il bisturi a metà longitudinalmente; si asciuga l'eccesso di EtOH e di H₂O.

Digestione dei campioni:

- Si mettono i campioni in termostato a 56°C (non in agitazione) per una notte;
- Si completa la digestione a 95°C per 8 min.

Recupero del DNA:

- Si centrifugano le provette per 10 min alla velocità di 17.000 rpm (giri per minuto), a temperatura ambiente, e si raccolgono circa 100 – 150 μ l di sovrantante, prestando attenzione a non aspirare Chelex o residui di campione;
- Si trasferiscono i campioni di DNA in soluzione in nuove “eppendorf” che devono essere congelate a - 20°C.

(2) Estrazione del DNA con fenolo-cloroformio da campioni di sangue

Preparazione delle soluzioni:

Tampone di lisi: TNE 1X (50mM TRIS-HCl pH 7,5; 10 mM NaCl; 5 mM EDTA):

1M TRIS-HCl pH 7,5: 10 ml
NaCl: 0,11 gr
200mM EDTA: 5 ml
in 200 ml di volume totale di ddH₂O

Preparazione dei campioni:

- Si preleva una piccola quantità di sangue che viene trasferita in provette “eppendorf” da 2 ml; il sangue viene lavato per due o più volte con l'aggiunta di 800 μ l di ddH₂O (l'acqua lisa le cellule e permette di estrarre l'emoglobina che altrimenti si lega al DNA e non permette una buona digestione); si centrifuga per 1 min; si elimina il sopranatante;
- Si aggiunge al pellet la soluzione di lisi:
850 μ l di TNE 1X

57 μ l di Proteinasi K (10 mg/ml)
83 μ l di SDS 10%

Digestione dei campioni:

- In agitazione a 57°C per una notte.

Estrazione del DNA:

Si trasferiscono i campioni digeriti in provette “eppendorf” con silicone, che separa la fase organica, contenente il fenolo, dalla fase acquosa, contenente il DNA. Si eseguono le seguenti estrazioni:

- 1^a estrazione con fenolo: si aggiunge al campione un volume uguale di fenolo; agitare delicatamente fino ad ottenere una miscelazione delle due fasi; centrifugare per 5 min a 10.000 rpm a temperatura ambiente; prelevare la fase superiore che viene trasferita in una nuova “eppendorf” con silicone;
- 2^a estrazione con fenolo: si ripete la prima estrazione; il campione pulito viene trasferito in provette “eppendorf” da 2 ml;
- 3^a estrazione con fenolo-cloroformio-alcool isoamilico: si aggiunge un volume uguale di fenolo-cloroformio-alcool isoamilico; agitare delicatamente fino ad ottenere una miscelazione delle due fasi; centrifugare per 1 min a 10.000 rpm a temperatura ambiente; la fase superiore viene trasferita in provette “eppendorf” da 2 ml;
- 4^a estrazione con cloroformio-alcool isoamilico: si aggiunge un volume uguale di cloroformio-alcool isoamilico; agitare delicatamente fino ad ottenere una miscelazione delle due fasi; centrifugare per 1 min a 10.000 rpm a temperatura ambiente; la fase superiore viene trasferita in provette “eppendorf” da 1,5 ml.

Purificazione del DNA mediante dialisi:

- In un becker si preparano 2 litri di TE (20 ml di TRIS 1M, pH 8,0; 4 ml di EDTA 0,5 M, pH 8,0; portare a 2 litri con ddH₂O); si preparano i tubi per dialisi, che vengono lavati con ddH₂O;
- La soluzione di DNA è trasferita nel tubo, che è chiuso e immerso nel becker con TE. La dialisi avviene in frigorifero per circa 20 - 24 ore.

(3) Estrazione del DNA con GUS

Preparazione delle soluzioni:

GUS, soluzione stock (7,05M guanidina tiocianato):

500 gr di GUS sciolti in 200 ml di ddH₂O

GUS, tampone di lisi (0,05M TRIS-HCl pH7; 0,025M EDTA pH8; 1,25% Triton X100; 4,23M GUS):

- 1M TRIS-HCl, pH 7,0: 12,5 ml
- 0,5M EDTA, pH 8,0: 12,5 ml
- TRITON X100: 3,125 ml
- ddH₂O: fino a 100 ml
- GUS soluzione stock: 150 ml (volume finale: 250 ml)
- GUS, soluzione di binding (0,05M TRIS-HCl pH7; 0,025M EDTA pH8; 4,23M GUS; diatomee 1%)
- 1M TRIS-HCl, pH 7,0: 12,5 ml
- 0,5M EDTA, pH 8,0: 12,5 ml
- ddH₂O: fino a 100 ml
- GUS soluzione stock: 150 ml
- Diatomee (Sigma): 2,5 gr
- volume finale: 250 ml
- GUS, soluzione di lavaggio (0,05M TRIS-HCl pH7,0; 4,23M GUS)
- 1M TRIS-HCl, pH 7,0: 25 ml
- ddH₂O: fino a 200 ml
- GUS soluzione stock: 300 ml
- volume finale: 500 ml
- TE (TRIS-HCl 10mM pH 8,0; EDTA 0,1 mM pH 8,0):
- 1M TRIS-HCl, pH 8,0: 1 ml
- 0,5 M EDTA, pH 8,0: 0,02 ml
- ddH₂O: fino a 100 ml di volume finale

Preparazione dei campioni:

- Si taglia un pezzetto di tessuto di circa 50 milligrammi (mg), che è trasferito in una provetta "eppendorf" da 1,5 ml contenente 500 ml di tampone di lisi GUS; si lavora con bisturi e pinze sterilizzate alla fiamma.

Digestione dei campioni:

- In rotazione a 57°C per una notte.

Recupero del DNA:

- Si centrifuga a temperatura ambiente per 10 minuti; si raccoglie il sopranatante;
- Si aggiungono al sopranatante 500 μ l di GUS soluzione di binding; in rotazione per 1 ora;
- Si centrifuga a temperatura ambiente per 1 minuto; si elimina il sopranatante.

Il DNA ora è legato ai microgranuli di silice in fondo alla provetta. Si effettuano due lavaggi del pellet, ciascuno con 500 μ l di GUS soluzione di lavaggio; si centrifuga a temperatura ambiente per 1 minuto; si elimina

il sovrantante. Si effettuano altri due lavaggi del pellet, ciascuno con 1 ml di EtOH 70%; si centrifuga a temperatura ambiente per 3 minuti; il pellet viene asciugato nelle “eppendorf” aperte in blocco termostatico a 56°C per 10 minuti.

Eluizione del DNA: si risospende il pellet in 200 μ l di TE per 15 min. a 56°C; si centrifuga a temperatura ambiente per 10 minuti; si trasferisce il sopranatante con il DNA in una nuova “eppendorf”.

I campioni di DNA si conservano in congelatore a - 20°C.

Controllo dell'estrazione del DNA

Il controllo dell'estrazione si effettua tramite elettroforesi di un'aliquota del DNA estratto in un minigel di agarosio (Fig. 27). Il DNA viene colorato immergendo il gel in una soluzione di bromuro d'etidio (EtBr), che si lega al DNA a filamento doppio e che emette luce quando viene esposto ai raggi ultravioletti di un transilluminatore. Il DNA integro si presenta come una singola banda compatta di alto peso molecolare. Il DNA degradato o digerito si presenta come una strisciata composta da una miriade di frammenti di peso molecolare variabile (Fig. 27). E' possibile quantificare la concentrazione del DNA presente in soluzione, tramite spettrofotometria, oppure tramite tecniche di PCR quantitativa.

Enzimi di restrizione ed analisi dei polimorfismi dei frammenti di restrizione (restriction fragment length polymorphism - RFLP)

Le sequenze di DNA possono essere determinate parzialmente ed indirettamente tramite le tecniche di RFLP. Il DNA totale viene estratto, tagliato (cioè digerito, o ristretto) con gli enzimi di restrizione. I frammenti originati dalla restrizione sono separati tramite elettroforesi in gel di agarosio, trasferiti tramite Southern blotting e fissati ad una membrana. Specifici frammenti di DNA sono poi individuati con sonde di DNA marcato (“probes”). Le sonde sono costituite da sequenze di DNA complementari ai frammenti che si intendono analizzare. Questo è il metodo utilizzato per l'analisi dei DNA fingerprinting multilocus, individuati tramite ibridazione con sonde multilocus (MLP).

Gli enzimi di restrizione sono proteine ad azione enzimatica che tagliano il DNA in punti specifici, caratterizzati da sequenze nucleotidiche specifiche (Fig. 28). Il primo enzima di restrizione fu identificato nel 1970, e fu denominato Hind II, dal nome latino del batterio *Hemophilus influenzae* da cui venne purificato. Gli enzimi di restrizione sono prodotti



Figura 27 - Elettroforesi del DNA in minigel di agarosio. Il DNA viene colorato immergendo il gel in una soluzione di bromuro d'etidio, che si lega al DNA a filamento doppio ed emette luce quando viene esposto ai raggi ultravioletti di un transilluminatore. Il DNA integro si presenta come una singola banda compatta di alto peso molecolare. Il DNA degradato o digerito si presenta come una strisciata composta da una miriade di frammenti di peso molecolare variabile.

Enzimi e siti di restrizione			
Eco RI	G AATTC	Hind III	A AGCTT
	CTTAA G		TTCGA A
Bam HI	G GATCC	Pst I	CTGCA G
	CCTAG G		G ACGTC
Hae II	PuGCGC Py	Taq I	T CGA
	Py CGCGPu		AGC T

Figura 28 - Gli enzimi di restrizione sono proteine ad azione enzimatica che tagliano il DNA in punti specifici, caratterizzati da sequenze nucleotidiche specifiche.

naturalmente ed utilizzati dai batteri per tagliare, inattivare ed eliminare il DNA estraneo (ad esempio DNA virale) che entra nella cellula. Gli enzimi di restrizione sono sistemi di difesa naturale che i batteri usano per difendersi dai DNA estranei che li invadono. A tutt'oggi sono stati isolati centinaia di enzimi di restrizione da più di 200 differenti ceppi batterici. I siti di restrizione sono palindromi, cioè l'ordine dei nucleotidi nel segmento di un filamento di DNA è l'inverso di quello nel filamento complementare. La sequenza di DNA può essere quindi letta in entrambi i sensi ed il sito di restrizione viene individuato in entrambi i filamenti della doppia elica. La lunghezza dei siti di restrizione varia, di solito, da 4 a 6 nucleotidi. Gli enzimi di restrizione sono utilizzati per digerire il DNA estratto dai campioni. Il DNA di alto peso molecolare estratto dai campioni viene posto in una soluzione che contiene un tampone appropriato per ottimizzare l'attività di un particolare enzima di restrizione alla temperatura ottimale. L'enzima di restrizione viene aggiunto alla soluzione e la reazione procede per qualche ora (di solito da 1 a 12 ore). L'enzima di restrizione legge il DNA, ed ogni volta che incontra il proprio sito di restrizione, lo taglia. Al termine della reazione il DNA è digerito, "ristretto". La soluzione non è più composta da molecole di DNA nativo a filamento lungo, ma da un insieme di gruppi di frammenti di DNA digerito, di piccole dimensioni. Le condizioni di reazione ottimali per la digestione variano a seconda dell'enzima di restrizione, e sono di solito indicate dal fornitore degli enzimi. I parametri critici per ottenere una buona digestione sono la temperatura e la concentrazione salina dei tamponi. E' quasi sempre possibile operare usando tre tipi di tampone: a bassa, media ed alta forza ionica (cioè 0 mM, 50 mM e 100 mM NaCl). I tamponi sono normalmente preparati ad una concentrazione 10x, e conservati congelati a - 20°C prima dell'uso. Un'unità di enzima di restrizione è definita come la quantità di enzima richiesta per digerire 1 μ g di DNA di fago *lambda* in un'ora. Tuttavia, nella pratica di laboratorio, per assicurare che le digestioni vengano completate, sia la concentrazione dell'enzima che il tempo di digestione sono incrementati. Le reazioni di solito contengono almeno il doppio della quantità di enzima teoricamente necessaria, e le digestioni continuano per 10 - 12 ore. Gli enzimi di restrizione sono conservati in tamponi contenenti glicerolo, a - 20°C, e devono essere scongelati il meno possibile per evitare l'inattivazione. Le soluzioni concentrate degli enzimi di restrizione devono essere diluite almeno 10 volte, per diluire il glicerolo che inibisce l'attività enzimatica. Al termine della digestione si rimuove un'aliquota di soluzione che viene utilizzata per il controllo di qualità della digestione su minigel di aga-

rosio. Il DNA digerito risulta composto da una serie di frammenti di dimensioni, e quindi di peso molecolare definito e discreto. Questi frammenti possono essere separati in elettroforesi su gel di agarosio e visualizzati tramite Southern blotting ed ibridazione con una sonda oligonucleotidica marcata. Le mutazioni nucleotidiche, ed anche le inserzioni/delezioni o le traslocazioni, possono modificare i siti di restrizione in particolari campioni. L'analisi dei frammenti di restrizione in una collezione di campioni può evidenziare variabilità genetica (Fig. 29).

L'RFLP è il metodo classico di analisi del DNA fingerprinting multi-locus. Nell'analisi di DNA fingerprinting è importante scegliere enzimi che abbiano siti di restrizioni solo all'esterno del repeat, per evitare che la digestione frammenti il repeat. Siti di restrizione ripetuti all'interno dei repeats potrebbero frammentare i repeats in segmenti troppo piccoli per essere individuabili.

Analisi dei frammenti di DNA tramite elettroforesi su gel di agarosio

Gli acidi nucleici sono carichi negativamente, quindi i frammenti di DNA migrano all'interno di un campo elettrico spostandosi verso il polo positivo (Fig. 30). La velocità di migrazione dei frammenti di DNA a filamento doppio è inversamente proporzionale al logaritmo del loro peso molecolare, ed è direttamente proporzionale al voltaggio applicato al sistema. L'elettroforesi avviene in gel porosi di agarosio o di acrilamide, che vengono immersi in una soluzione di elettroliti. I gel hanno, ad un'estremità, una fila di pozzetti che sono usati per il caricamento dei campioni di DNA. Quando viene erogata corrente elettrica al sistema, i frammenti di DNA entrano nel gel e si spostano verso il polo positivo, seguendo una direzione rigorosamente parallela al campo elettrico. Al termine dell'elettroforesi, dopo un periodo di tempo che può essere di pochi minuti (per i minigel di controllo della PCR) fino ad alcuni giorni (per i DNA fingerprinting), i frammenti di DNA risultano separati in gruppi costituiti da frammenti di peso molecolare omogeneo. I frammenti di alto peso molecolare sono localizzati a poca distanza dalla linea di partenza. Al contrario, i frammenti di basso peso molecolare sono localizzati verso l'estremità opposta del gel. La concentrazione di agarosio o di acrilamide influenza la velocità di migrazione, ma soprattutto la focalizzazione e quindi la risoluzione dei frammenti, in relazione al loro peso molecolare. Di solito si usano gel con concentrazione dell'1 - 2% di agarosio. Il gel è colorato per immersione in una soluzione di EtBr, che si lega esclusivamente al DNA a doppia elica. L'EtBr è fluorescente alla

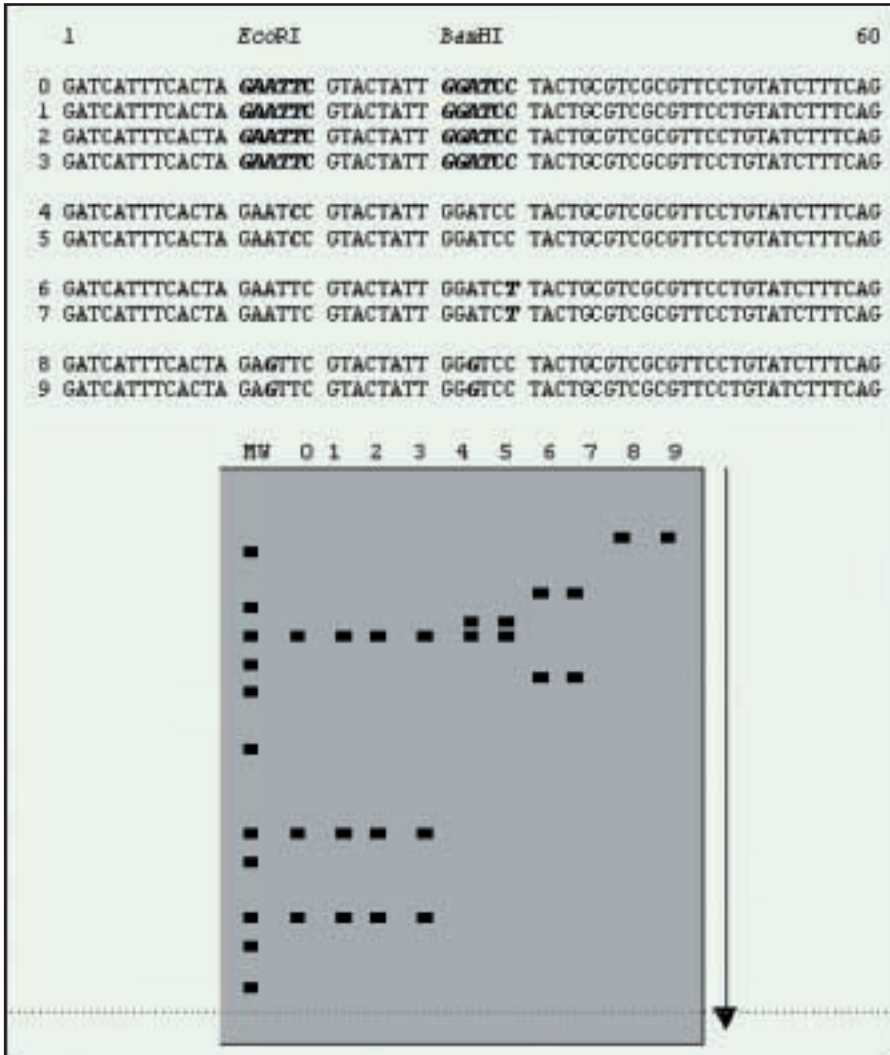


Figura 29 - Il DNA digerito con gli enzimi di restrizione risulta composto da una serie di frammenti di dimensioni discrete e quindi di peso molecolare definito. Questi frammenti possono essere separati in elettroforesi su gel di agarosio e visualizzati tramite Southern blotting ed ibridazione con una sonda oligonucleotidica marcata. Le mutazioni nucleotidiche, ed anche le inserzioni/delezioni o le traslocazioni, possono modificare i siti di restrizione in particolari campioni. L'analisi dei frammenti di restrizione in una collezione di campioni può evidenziare variabilità genetica.

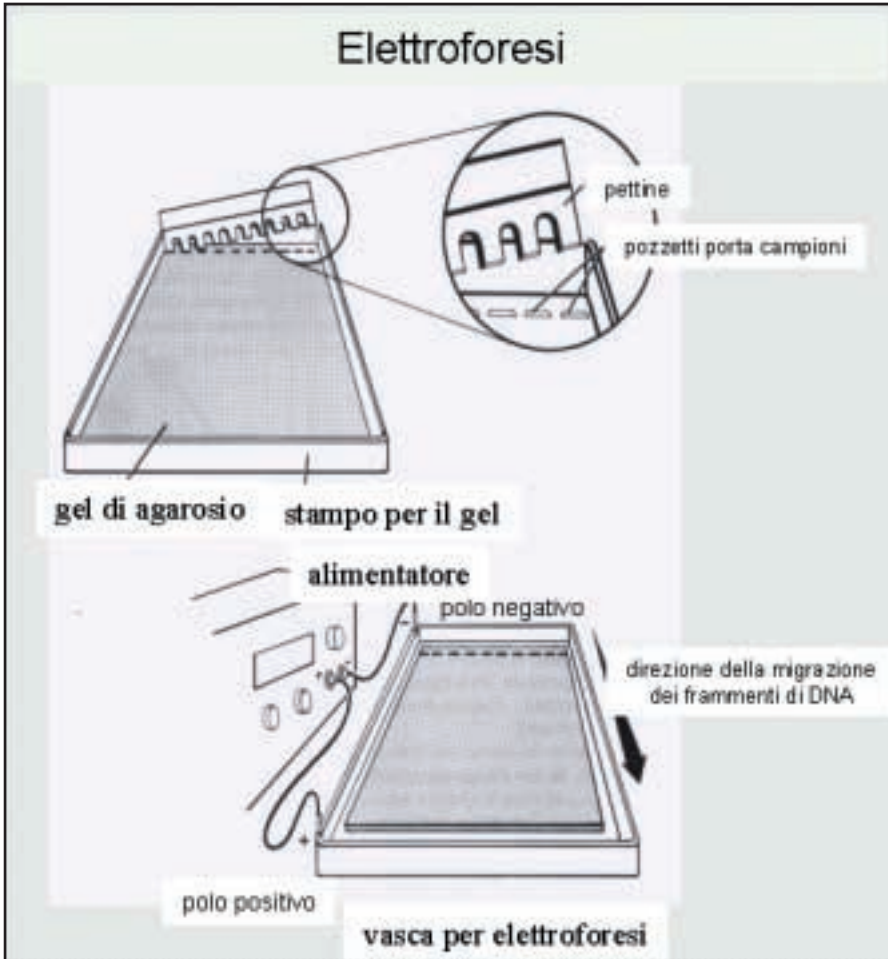


Figura 30 - Schema di un'apparato per l'elettroforesi del DNA in gel di agarosio. Gli acidi nucleici sono carichi negativamente, quindi i frammenti di DNA migrano all'interno di un campo elettrico spostandosi verso il polo positivo. La velocità di migrazione dei frammenti di DNA a filamento doppio è inversamente proporzionale al logaritmo del loro peso molecolare, ed è direttamente proporzionale al voltaggio applicato al sistema. L'elettroforesi avviene in gel di agarosio o di acrilamide, che vengono immersi in una soluzione di elettroliti. I gel hanno, ad un'estremità, una fila di pozzetti che sono usati per il caricamento dei campioni di DNA. Quando viene erogata corrente elettrica, i frammenti di DNA entrano nel gel e si spostano verso il polo positivo, seguendo una direzione rigorosamente parallela al campo elettrico.

luce ultravioletta e quindi consente di evidenziare il DNA nel gel. Il DNA appare come una banda fluorescente compatta se il campione non è degradato, o come una strisciata se il campione è degradato (Fig. 27). Il DNA molto degradato può essere non visibile nei minigel di agarosio. È importante notare che l'EtBr evidenzia tutto il DNA presente nel minigel, cioè il DNA del campione più l'eventuale DNA contaminante. L'elettroforesi in gel di agarosio viene utilizzata per il controllo dell'estrazione, della digestione e dell'amplificazione del DNA, oltre che per le analisi degli RFLP. Le analisi dei microsatelliti e delle sequenze nucleotidiche vengono effettuate utilizzando gel di acrilamide.

Southern blotting

I gel di agarosio sono fragili e difficilmente potrebbero essere manipolati per effettuare l'ibridazione, i lavaggi e l'autoradiografia. Inoltre, i frammenti di DNA tendono a diffondere rapidamente nell'agarosio quando, alla fine dell'elettroforesi, la corrente elettrica viene interrotta. Questi problemi vengono risolti tramite la tecnica del Southern blotting (Southern 1975), che consiste nel trasferimento del DNA dal gel di agarosio ad una membrana di nylon. La procedura del Southern blotting è descritta dalla figura 31. Il DNA a doppio filamento viene denaturato immergendo il gel di agarosio in una soluzione alcalina, poi il gel viene neutralizzato in una soluzione ad alta concentrazione salina, per mantenere il DNA denaturato. Si depone una membrana di nylon sul gel e si prepara il blotting. Gli strati di carta assorbente asciutta che sono posti sopra la membrana attraggono per capillarità il tampone. Questo flusso verticale di tampone trascina con sé il DNA che esce dal gel ed entra nella membrana. Le membrane sono meccanicamente robuste, chimicamente resistenti ed hanno carica netta positiva che facilita l'assorbimento ed il legame dei frammenti di DNA, che sono carichi negativamente. Al termine del blotting, il DNA viene immobilizzato sulla membrana tramite esposizione ai raggi UV per pochi minuti in un transilluminatore. Gli UV interrompono la continuità delle catene di DNA, dissociando i legami chimici fra nucleotidi ed esponendo atomi carichi negativamente che legano insolubilmente i frammenti di DNA alla membrana.

Ibridazione molecolare

Al termine del Southern blotting la membrana, che contiene i frammenti di DNA fissati, è utilizzabile per l'ibridazione con una sonda

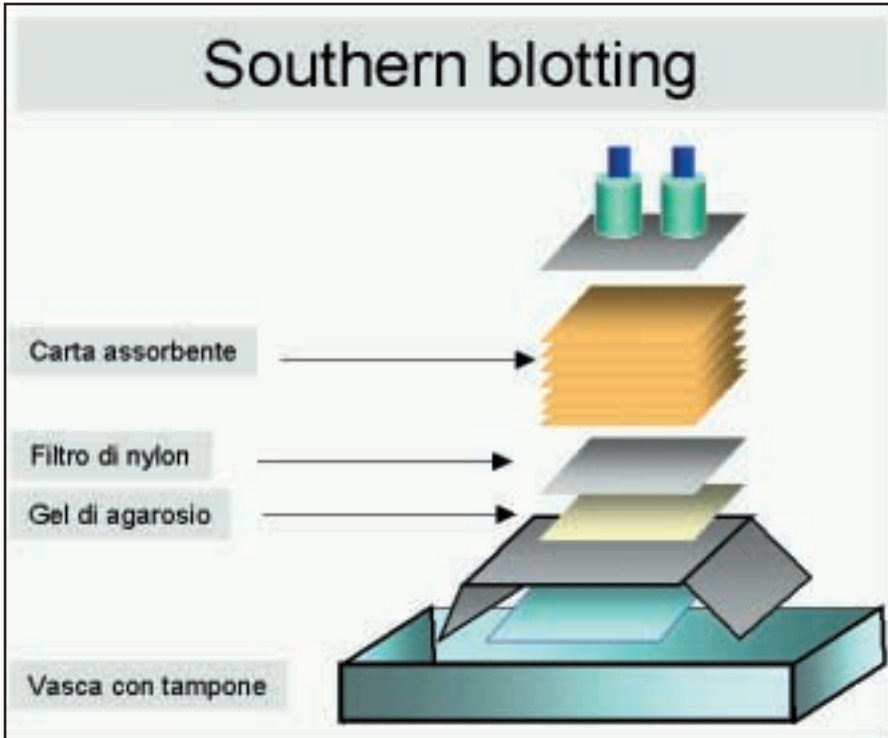


Figura 31 - Southern blotting. Consente il trasferimento dei frammenti di DNA, separati tramite elettroforesi in gel di agarosio, ad una membrana per blotting. Si prepara una vasca contenente una apposita soluzione tampone. Si taglia un foglio di carta assorbente che ha le dimensioni del gel di agarosio e che viene immersa nel tampone da entrambi i lati. Si prepara una membrana per blotting della dimensione del gel. Si depone il gel sulla carta assorbente bagnata e si depone la membrana bagnata sul gel. Si pone uno strato di molti fogli di carta assorbente asciutta sulla membrana. Tutto il sistema è mantenuto compresso tramite alcuni pesi. Il flusso capillare di tampone dalla vasca ai fogli di carta assorbente determina il trasferimento (blotting) dei frammenti di DNA dal gel di agarosio alla membrana.

oligonucleotidica (Fig. 32). L'ibridazione consiste nel provocare l'appaiamento fra la sonda oligonucleotidica marcata e le sequenze complementari del DNA dei campioni immobilizzate sulla membrana. Lo scopo è quello di ottenere un segnale, radioattivo oppure luminoso, in corrispondenza e solo in corrispondenza del punto in cui sono localizzati i frammenti di DNA che si vogliono evidenziare. Il segnale luminoso viene evidenziato esponendo la membrana ibridata ad una lastra fotosensibile (autoradiografia). Esistono diversi protocolli di ibridazione che sono calibrati per ottimizzare il segnale in relazione alle sonde utilizzate. Occorre

standardizzare le condizioni di ibridazione per ogni tipo di sonda. Presso il Laboratorio di Genetica dell'INFS si utilizza il seguente protocollo.

Il protocollo INFS per le analisi di DNA fingerprinting multilocus

(1) Estrazione del DNA con fenolo-cloroformio da campioni di sangue

Per le analisi di DNA fingerprinting multilocus si usano, di solito, campioni di sangue. Il DNA viene estratto col metodo del fenolo-cloroformio, seguendo il protocollo illustrato sopra.

(2) Quantificazione del DNA allo spettrofotometro

Le estrazioni di DNA vengono controllate tramite minigel di agarosio, per verificare che la qualità del DNA sia buona, cioè che il DNA sia

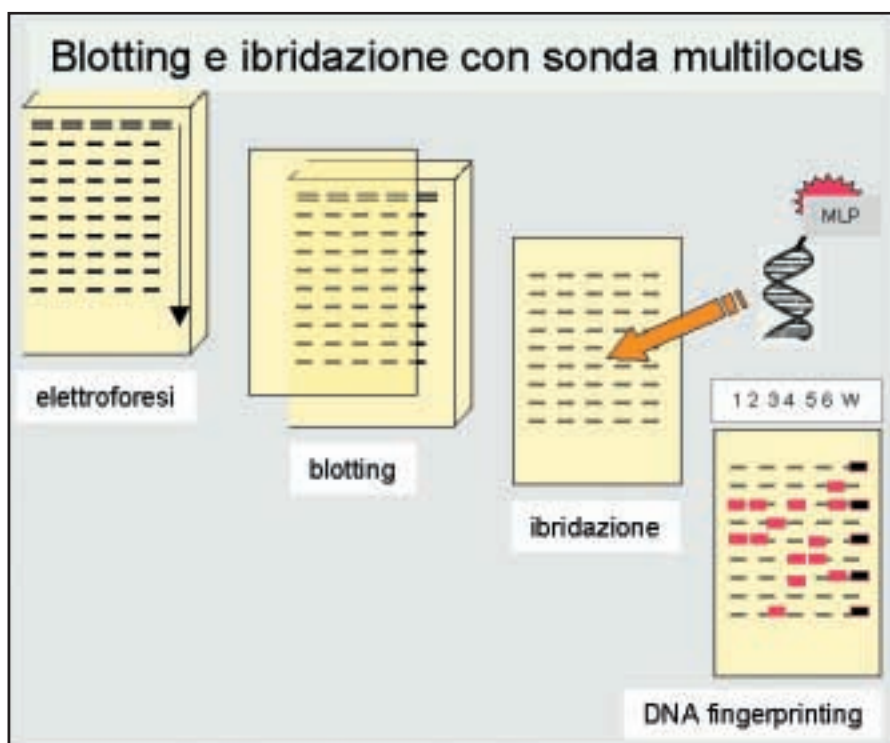


Figura 32 - Ibridazione con sonde multilocus. Una sonda multilocus marcata radioattivamente, o con marcature chimiche, si lega alle sequenze di DNA complementari e ne consente la visualizzazione.

integro. Poi si controlla la concentrazione del DNA tramite analisi allo spettrofotometro. Per ottenere autoradiografie di buona qualità occorre mettere in digestione circa 10 - 12 μg ($1 \mu\text{g} = 1 \gamma$; gamma) di DNA per ogni campione. Le concentrazioni spettrofotometriche del DNA estratto devono essere di circa 50 (valore minimo) - 300 (valore massimo) $\mu\text{g}/\text{ml}$.

(3) Digestione con gli enzimi di restrizione

Per assicurare che la digestione sia completa, per ogni campione vengono fatte due digestioni consecutive. La prima digestione viene fatta alla temperatura di 37°C per una notte, in un volume totale di 200 μl , che includono:

- non più di 175 μl di DNA in soluzione, corrispondenti a circa 10 - 12 γ di DNA;
- 3 - 5 μl di enzima di restrizione;
- 20 μl di tampone di digestione;
- eventuale aggiunta di ddH_2O sterile per portare a volume desiderato.

La seconda digestione viene fatta alla temperatura di 37°C per circa 4 - 5 ore, aggiungendo per ogni campione 100 μl di soluzione di digestione (1 μl di enzima di restrizione; 10 μl di tampone di digestione; 89 μl di ddH_2O).

Al termine della seconda digestione si controlla su minigel che il DNA sia ben digerito, cioè che non siano presenti frammenti definiti di alto peso molecolare, ma abbia l'aspetto di una strisciata omogenea (Fig. 27). Il controllo della digestione viene fatto caricando 5 μl di DNA digerito + 5 μl di EtBr in un minigel di agarosio, che corre in elettroforesi a 150 - 200 volts per pochi min. Il DNA digerito può essere conservato in congelatore a -20°C .

(4) Precipitazione e risospensione del DNA digerito

Al DNA digerito si aggiungono circa 30 μl di sodio acetato (NaAc) 3M (cioè circa 1/10 del volume della soluzione di DNA digerito), meglio se freddo. Poi:

- si aggiungono circa 2,2 volumi di etanolo 100% (circa 700 μl), meglio se freddo;
- si agita delicatamente e si ripone la soluzione a -20°C per 30 min;
- si centrifuga a 17.000 rpm per 30 min a 4°C ; in questo modo il DNA precipita come pellet;
- si elimina il soprannatante senza staccare pellet;
- si aggiunge 1 ml di etanolo 70%;
- si centrifuga a 17.000 rpm per 15 min a 4°C ;

- si elimina il sopranatante senza staccare pellet;
- si asciugano i campioni a temperatura ambiente;
- si risospende il pellet di DNA in 20 - 25 μ l di TE.

(5) Elettroforesi su gel di agarosio

Si prepara un gel di agarosio concentrato all'1% oppure allo 0,8% in tampone TBE; si aggiungono 5 μ l di blu di bromofenolo + glicerolo al 30% in ogni campione (il blu forma una linea ben visibile che migra più velocemente di ogni frammento di DNA e serve per controllare il progresso dell'elettroforesi; il glicerolo serve per addensare i campioni e facilitare il loro caricamento nei pozzetti del gel di agarosio).

Organizzazione dei campioni sul gel. L'ordine con cui si caricano i campioni sul gel è molto importante, perchè l'analisi dei risultati, cioè la valutazione del DNA fingerprinting, viene effettuata comparando direttamente le velocità di migrazione elettroforetica dei vari frammenti di DNA fra i diversi campioni. Quando si effettuano test di parentela, viene usato il seguente schema di caricamento:

4. 1 = standard di peso molecolare (DNA di *lambda* digerito con Hind III, o altri pesi molecolari)
5. 2 = campione di controllo
6. 3 = padre presunto
7. 4 = figlio 1
8. 5 = figlio 2
9. 6 = madre presunta
10. 7 = peso molecolare

La corsa elettroforetica viene effettuata a 25 - 30 volts per 2 - 3 giorni. Il tampone TBE 1X delle vasche viene completamente sostituito ogni 24 ore.

(6) Southern blotting

Al termine dell'elettroforesi i frammenti di DNA sul gel di agarosio vengono denaturati tramite una serie di bagni. Il primo bagno viene fatto in una soluzione depurinante (27,33 ml/lit HCl) per 15 min. su un piano agitatore. La depurinazione rompe i legami tra purina e purina, e favorisce il passaggio dei frammenti di DNA dal gel alla membrana. Il gel viene lavato in ddH₂O. Poi il gel viene immerso in soluzione denaturante (0,5 M NaOH = 20 gr/lit; 1,5 M NaCl = 88gr/lit) per 15 min per 2 volte. Il gel viene ancora lavato in ddH₂O, poi viene immerso in soluzione neutralizzante (0,5 M Tris = 60,55gr/lit; 1,5 M NaCl = 88 gr/lit; EDTA Na = 0,4 gr/lit; pH 7,2), per 15 min per 2 volte.

Il blotting, cioè il trasferimento per capillarità dei frammenti di DNA dal gel di agarosio alla membrana di nylon, viene effettuato in una soluzione di blottaggio (20x SSC: 3 M NaCl = 175,33 gr/lt; 0,5 M Na citrato tribasico biidrato 88,23 gr/lt; pH 7,0 - 7,5). Poi si prepara l'apparato per il Southern blotting (Fig. 31), che avviene nel corso di una notte.

(7) Preibridazione ed ibridazione

Al termine del blotting, la membrana viene preparata per l'ibridazione con la sonda marcata. La membrana viene lavata in 5X SSC (50 ml in 200 ml), poi il DNA viene fissato tramite esposizione per 3 min agli UV sul transilluminatore. La membrana viene poi lavata in 1X SSC (10 ml in 200 ml di ddH₂O) a 50°C in rotazione.

- Preibridazione. La membrana viene preibridata 1 volta per 20 min a 50°C in rotazione, in 50 ml di soluzione di preibridazione (49,5 ml di 0,5 M Na₂HPO₄; pH 7,2 = 71 gr/lt; 0,5 ml SDS 10% = 100 gr/lt).
- Ibridazione. La membrana viene preibridata 1 volta per 20 min a 50°C in rotazione, in 20 ml di soluzione di ibridazione, composta da:
 - 18 ml di soluzione di preibridazione
 - 2 ml di block solution (10 gr di caseina in 100 ml di soluzione di lavaggio # 2)
 - a cui si aggiungono 10 ml di sonda marcata.

Al termine dell'ibridazione la membrana viene ripetutamente lavata:

- 1° lavaggio: 2 volte per 10 min a 50°C, in rotazione, ogni volta con 100 ml di soluzione di lavaggio # 1:
 - 166 ml ddH₂O
 - 32 ml 0,5 M Na₂HPO₄; pH 7,2 (= 71 gr/lt)
 - 2 ml SDS 10%
- 2° lavaggio: 2 volte per 10 min a temperatura ambiente, in piano agitatore, ogni volta con 200 ml di soluzione di lavaggio # 2:
 - 100 ml di soluzione di lavaggio # 2 4X (acido maleico: 55,2 gr/lt; NaCl 34,8 gr/lt; pH 7,5 con NaOH) in 400 ml di ddH₂O.

(8) Incubazione con rivelatore. Esposizione e sviluppo delle lastre autoradiografiche

Il protocollo di DNA fingerprinting INFS utilizza le due sonde multi-locus di Alec Jeffreys, e cioè la sonda 33.6, che ha un periodo di incubazione di circa 18 ore nell'uomo e nei primati, e di circa 2-3 giorni in altri mammiferi e negli uccelli; la sonda 33.15, che ha un periodo di incuba-

zione di 3 ore nell'uomo e nei primati, e di circa 1 giorno in altri mammiferi e negli uccelli. Queste due sonde sono protette da brevetto (UK Patent No. 2166445), e sono commercializzate dalla Cellmark Diagnostic (PO Box 265, Abingdon, Oxon OX14 IYX, UK; cellmark@orchidbio.co.uk). Le sonde Cellmark sono marcate con l'enzima fosfatasi alcalina e commercializzate come "NICE™ non-isotopic probe system", un prodotto che consente di evitare l'uso di isotopi radioattivi, di evitare di marcare le sonde in laboratorio, e che assicura risultati di elevata qualità. L'attività della fosfatasi alcalina viene evidenziata applicando alla membrana il reagente Lumi-Phos™ o il più recente e CDP-Star™. Questi reagenti attivano l'enzima fosfatasi alcalina, che avvia una reazione chemiluminescente. L'emissione di luce continua per almeno 5 giorni e consente di esporre rapidamente lastre fotografiche per raggi X. In alternativa al sistema NICE™, è possibile clonare, purificare e marcare le sonde nel proprio laboratorio. Le sonde oligonucleotidiche possono essere marcate con isotopi radioattivi, ad esempio, legando atomi di fosforo 32 ad un nucleotide (ad esempio: α -³²PdCTP), oppure incorporando nelle sonde marcatori non-radioattivi (ad esempio, fosfatasi alcalina).

Al termine dei lavaggi la membrana viene cosparsa di Lumi-Phos™ o di CDP-Star™, chiusa in un sacchetto di plastica, e collocata in una cassetta per autoradiografia. A questo punto si entra in camera oscura e si pone una lastra per autoradiografia a contatto con la membrana. Si chiude la cassetta per autoradiografia e si lascia la lastra in esposizione per tutto il tempo necessario. La lastra autoradiografica viene sviluppata tramite bagni in soluzioni di sviluppo e di fissaggio. La lastra sviluppata viene lavata con acqua e viene lasciata ad asciugare.

Se necessario, la membrana può essere reibridata, cioè la sonda fissata ai frammenti di DNA complementari può essere dilavata con SDS 0,1% (0,5 ml SDS; 49,5 ml H₂O) per 15 min a 80°C in rotazione, e poi con 1x SSC (10 ml in 200 ml) a 50°C in rotazione. A questo punto la membrana può essere reibridata, ripetendo le procedure di preibridazione, ibridazione, lavaggio, ecc., utilizzando la stessa od un'altra sonda.

L'esecuzione completa del protocollo DNA fingerprinting multilocus INFS richiede da 7 a 9 giorni lavorativi.

Struttura delle sonde multilocus usate in genetica forense

Le sonde ("probes") utilizzate in genetica forense sono costituite da DNA o RNA a filamento singolo o doppio che contengono le sequenze complementari alle sequenze ripetute dei loci usati per ottenere i DNA

fingerprinting. Questi marcatori genetici sono stati scelti perché sono molto polimorfici, e perché possono essere tipizzati in quasi tutte le specie di vertebrati, ibridizzando le sonde multilocus al DNA dei campioni in condizioni di bassa stringenza. I minisatelliti sono regioni del DNA genomico composte da brevi sequenze ripetute in tandem. Il polimorfismo deriva dalle differenze di peso molecolare fra i diversi alleli. Le differenze di peso molecolare derivano del diverso numero di repeat che compongono gli alleli. Ogni locus minisatellite presenta molteplici alleli in ogni popolazione, e quindi i valori di eterozigosi individuale sono molto elevati. La variabilità fra alleli deriva probabilmente da crossing-over asimmetrico o da DNA slippage durante le meiosi che portano alla gametogenesi. La variabilità nella lunghezza degli alleli ai loci minisatellite può essere evidenziata digerendo il DNA con enzimi di restrizione, che tagliano ai fianchi, ma non all'interno, dei minisatelliti. Sono stati identificati numerosi loci minisatellite, che mappano all'interno oppure in prossimità di geni come il gene per l'insulina umana, alcuni geni per le globine e certi oncogeni. Il primo minisatellite è stato scoperto e caratterizzato da Alec Jeffreys e collaboratori nel 1985. Questo minisatellite, che venne identificato nel primo introne del gene della mioglobina umana (Fig. 14), è composto da quattro repeat di una sequenza lunga 33 nucleotidi. Il minisatellite è fiancheggiato da sequenze non-ripetute che contengono due siti di restrizione per l'endonucleasi HinfI. La digestione con HinfI consente di isolare un segmento dell'introne lungo 169 nucleotidi, che contiene tutto il minisatellite. Digerendo questo segmento con altre endonucleasi che hanno siti di restrizione all'interno del repeat, è stato possibile isolare e clonare il monomero lungo 33 nucleotidi. Questa sequenza (detta sequenza "core") è stata utilizzata per costruire la sonda multilocus 33.7. Clonando loci minisatelliti identificati tramite la sonda 33.7 è stato possibile identificare nuovi loci, ciascuno dei quali conteneva da 3 a 29 copie di un repeat identico o simile alla sequenza core del minisatellite contenuto nel primo introne del gene della mioglobina umana. Ciascuno di questi loci mappa in una regione unica del genoma umano. Da due di questi loci sono state ricavate le sonde multilocus 33.15 e 33.6, che sono ipervariabili, cioè presentano alleli multipli in popolazioni umane e di molti altre specie di vertebrati.

Le sonde multilocus (MLP) di Jeffreys:

- La sonda 33.15 è composta da 32 nucleotidi che sono suddivisi in due sequenze ripetute ciascuna di 16 nucleotidi:



Questa sonda è complementare al repeat AGAGGTGGGCAG-GTGG.

- La sonda 33.6 è composta da 37 nucleotidi che includono tre sequenze ripetute ciascuna di 11 nucleotidi:



Questa sonda è complementare al repeat AGGGCTGGAGG.

I nucleotidi sottolineati corrispondono alla sequenza “core”: GGGCAGGAXG, che è presente, anche se non completamente conservata, in tutti i repeat di ciascuno dei minisatelliti individuati dalle sonde di Jeffreys.

Ogni lastra autoradiografica o fluorografica identifica profili individuali che sono normalmente composti da 10 a 30 frammenti di DNA, ognuno dei quali rappresenta un'allele. Sebbene le sonde di Jeffreys siano state sviluppate ed utilizzate principalmente per l'analisi del DNA umano, esse sono utili anche per l'analisi del DNA fingerprinting in altre specie animali. Buoni DNA fingerprinting sono stati ottenuti in molte specie di vertebrati inclusi canidi, felidi, uccelli, pesci. Sono state clonate altre sonde che sono utilizzabili in genetica forense umana ed anche in specie animali e vegetali, come, ad esempio, il minisatellite dell' α -globina ed il repeat contenuto nel batteriofago M13. Si utilizzano anche sonde oligonucleotidiche sintetizzate chimicamente.

Il tasso di mutazione dei minisatelliti individuati da Jeffreys è stimato in circa 1×10^{-4} nuovi alleli per gamete per generazione, che significa che ogni allele muta una volta ogni 10.000 gameti. Questo tasso di mutazione è più grande di almeno un ordine di grandezza del tasso di ricombinazione simmetrica, ed è più grande di alcuni ordini di grandezza del tasso di sostituzione nucleotidica a loci selettivamente neutrali.

Interpretazione dei DNA fingerprinting

L'analisi delle lastre autoradiografiche è complicata a causa del gran numero di frammenti (alleli) che sono presenti nei profili individuali e della grande variabilità inter-individuale dei DNA fingerprinting. Il problema è di identificare quali frammenti di ogni profilo sono “identici” e quali sono “differenti” da un individuo all'altro. La determinazione delle corrispondenze fra i frammenti individuali può essere fatta visivamente, oppure utilizzando sistemi computerizzati di analisi dell'immagine. In ogni caso occorre tener conto dei fattori sperimentali che possono causare imprecisione nell'identificazione dei profili di DNA fingerprinting, quali l'ampiezza e l'intensità dei frammenti, la qualità della risoluzione elettroforetica, le variazioni di mobilità elettroforetica all'interno di un gel, le

eventuali deformazioni della migrazione elettroforetica in zone diverse del gel. Questi errori sperimentali fanno sì che due frammenti identici possano essere visualizzati come diversi, oppure, viceversa, che due frammenti di peso molecolare diverso siano visualizzati come identici. Occorre perciò stabilire dei criteri di identificazione della mobilità, e quindi del peso molecolare dei frammenti, che tengano conto degli errori sperimentali. Di solito due frammenti sono identificati come identici se hanno un peso molecolare che è determinato entro 3 unità di deviazione standard in ogni senso di migrazione. In sistemi elettroforetici ben calibrati, ogni frammento ha una deviazione standard corrispondente a circa lo 0.6% del suo peso molecolare. Per esempio, la definizione di un frammento di 4.000 nucleotidi può variare di 25 nucleotidi x 6, e quindi frammenti di 4.000 +/- 150 nucleotidi sono assegnati allo stesso allele e sono considerati identici.

Oltre ai problemi tecnici, che derivano dalla qualità delle procedure di laboratorio e dalla identificazione dei frammenti nei DNA fingerprinting, i sistemi MLP presentano altri problemi di applicazione. Normalmente la variabilità dei sistemi MLP è così elevata che è estremamente improbabile, ma non è impossibile, che due individui abbiano lo stesso DNA fingerprinting. E' difficile stimare quali siano le probabilità di identità (PID) fra due individui presi a caso nella popolazione di riferimento, poiché le basi genetiche (cioè: quanti loci sono stati evidenziati, quanti alleli per ogni locus, quanti loci sono omozigoti oppure eterozigoti nel campione analizzato, ecc.) dei frammenti che compongono un DNA fingerprinting non sono note. Non è semplice stabilire le relazioni di linkage fra gli "alleli" di un sistema MLP. Se un sottoinsieme di "alleli" MLP sono linked, allora sono ereditati in blocco dai genitori, e quindi il potere di identificazione può essere molto minore di quanto potrebbe apparire se tutto il sistema fosse composto da "alleli" non linked.

Le sonde SLP individuano singoli specifici loci minisatellite. Ogni individuo risulta pertanto caratterizzato ad ogni locus da uno (se è omozigote) o da due (se è eterozigote) frammenti di DNA, e non da un sistema di bande multiple come nei sistemi MLP. Questi loci VNTR hanno molti alleli e presentano valori di eterozigosità molto elevati. Il genotipo individuale multilocus viene determinato analizzando separatamente più loci (di solito 4 - 6) e cumulando i dati, cioè formando genotipi multilocus che corrispondono ai DNA fingerprinting dei sistemi MLP. I loci VNTR hanno il vantaggio di essere ben identificati di avere un massimo di due alleli per individuo, ma molti alleli nelle popolazioni.

Amplificazione del DNA

L'analisi dei frammenti di restrizione tramite Southern blotting richiede da 1 a 10 μ g di DNA integro, di alto peso molecolare. Tuttavia, spesso i campioni biologici disponibili per le analisi di genetica forense contengono poco DNA, talvolta degradato. L'amplificazione del DNA diventa quindi una procedura necessaria per ottenere la quantità di campione sufficiente per le analisi molecolari. Sequenze di DNA composte da poche decine a poche migliaia di nucleotidi possono essere amplificate efficacemente utilizzando la reazione di polimerizzazione a catena (Polymerase Chain Reaction - PCR). La PCR consente di sintetizzare in provetta alcuni microgrammi di DNA a partire da pochi picogrammi di campione. In teoria è possibile utilizzare una singola molecola integra di DNA target (per DNA target si intende il gene o la sequenza di DNA che si deve analizzare e che, quindi, deve essere amplificata). Sequenze presenti in copia singola nel DNA dei campioni possono essere amplificate fino a 10 milioni di volte in poche ore. Il DNA amplificato è costituito da copie multiple della sequenza target, ed è sufficientemente puro da poter essere direttamente sequenziato oppure analizzato tramite altre tecniche molecolari. La PCR (Fig. 33) avviene ricostruendo in provetta le condizioni chimiche necessarie per realizzare una sintesi di DNA. Per prima cosa occorre identificare il gene o la sequenza di DNA che si desidera amplificare (per esempio, la regione di controllo del DNA mitocondriale, oppure un locus microsatellite). La sequenza da amplificare è fiancheggiata da entrambe le parti da sequenze che devono essere almeno parzialmente note. Infatti, per avviare la PCR occorre sintetizzare in laboratorio una coppia di oligonucleotidi ("primer") che siano almeno parzialmente complementari alle sequenze fiancheggianti. Gli oligonucleotidi utilizzati come primer sono sintetizzati chimicamente, e di solito sono prodotti da laboratori commerciali specializzati che utilizzano macchine automatiche. Occorre fornire al laboratorio commerciale le sequenze dei primer che devono essere determinate accuratamente. I primer si agganciano alle sequenze complementari fiancheggianti e consentono l'avvio del processo di duplicazione della sequenza target. Ogni PCR consiste di un ciclo, ripetuto molte volte, composto dai seguenti passaggi: denaturazione del DNA del campione; aggancio dei primer alle sequenze fiancheggianti; estensione dei primer tramite l'azione enzimatica di una DNA polimerasi, che si conclude con la replicazione completa di entrambi i filamenti della sequenza target. La PCR avviene in una provetta che contiene: il DNA del campione, i due primer, l'enzima DNA polimerasi, una certa

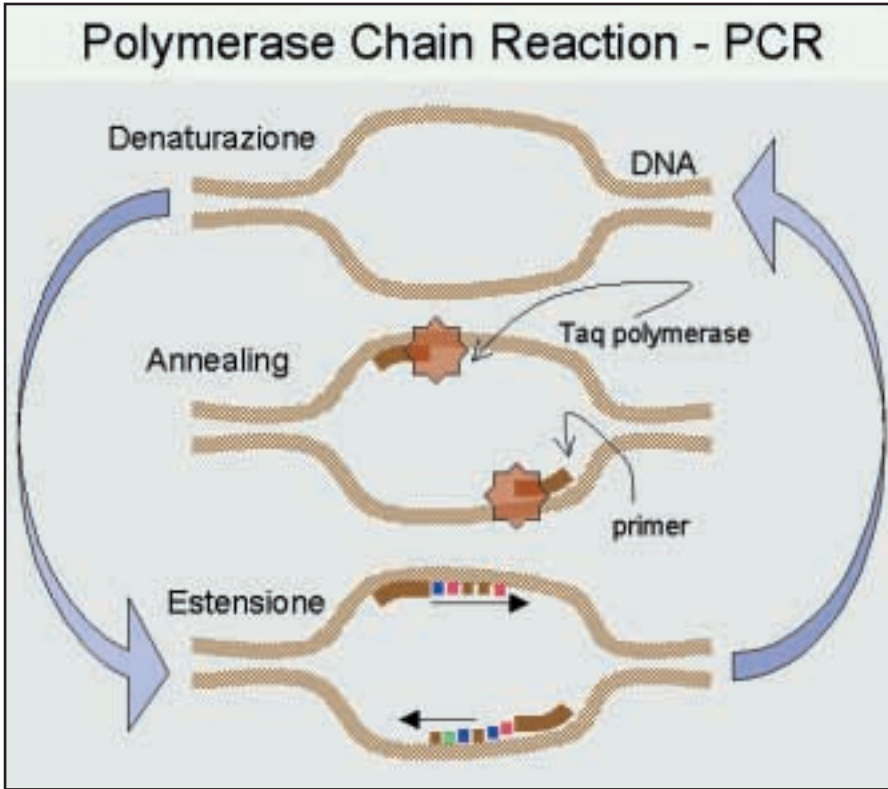


Figura 33 - La reazione di polimerizzazione a catena del DNA (Polymerase Chain Reaction; PCR).

quantità di nucleotidi liberi, il tutto in soluzione in un tampone che ottimizza le condizioni di sintesi del DNA. Ogni provetta per PCR è sistemata in un termostato (termocicizzatore) che esegue, con grande precisione, un ciclo termico prefissato, passando rapidamente dalla temperatura di denaturazione (superiore ai 90°C, per pochi secondi), alla temperatura di aggancio dei primer (variabile da circa 40 a 55°C, per pochissimi secondi), alla temperatura di estensione (72°C per alcuni secondi), per poi ritornare di nuovo alla temperatura di denaturazione, e così via per 20 - 40 volte. I termocicizzatori sono costruiti in modo da ospitare 96 provette da 0.2 ml, oppure piastre di tipo "microtiter" a 96 pozzetti, entro cui si miscelano i reagenti per le PCR. Poichè il ciclo termico può essere realizzato rapidamente, è possibile amplificare fino a 96 campioni in una o due ore. Inizialmente il DNA del campione è

denaturato per via termica, cioè innalzando la temperatura della reazione fino a 90° - 95°C. Quando il DNA è denaturato, cioè la doppia elica è completamente dissociata nei due filamenti singoli, ognuno dei primer si aggancia (“annealing”) alla sequenza complementare che fiancheggia la sequenza di DNA target. La temperatura di annealing dipende dalla lunghezza dei primer e dalla loro sequenza. Di solito varia fra i 40 e i 55°C, e deve essere determinata e controllata attentamente per evitare che i primer non si aggancino, oppure che si aggancino a sequenze non-target, producendo così amplificazioni aspecifiche. Quando l’annealing è completato, interviene una DNA polimerasi termoresistente (“Taq polymerase”), che avvia la duplicazione del DNA target a partire dai siti di priming. La reazione di polimerizzazione avviene in una soluzione tampone che mantiene un pH ottimale, una concentrazione ottimale di ioni Mg⁺⁺, e che contiene nucleotidi liberi in forma attiva (dATP, dCTP, dGTP e dTTP) che possono essere incorporati nel corso della sintesi di DNA. La sintesi avviene per estensione di entrambi i primer. La Taq polymerase catalizza l’estensione dei primer e produce due nuovi filamenti di DNA, i quali sono complementari ai filamenti della sequenza target. L’estensione dei primer avviene ad una temperatura ottimale per la sintesi del DNA, di solito a 72°C. La PCR procede quindi grazie ad un ciclo di tre temperature: la temperatura di denaturazione del DNA, la temperatura di annealing dei primer alla sequenza target (40°C - 55°C), la temperatura di estensione dei primer (72°C). I tempi di permanenza della reazione di PCR alle tre temperature possono variare, ma di solito sono piuttosto brevi, di pochi secondi. Al termine del primo ciclo, ogni copia della sequenza target presente nel campione è stata replicata una prima volta. A questo punto il ciclo viene ripetuto una seconda volta. Al termine del secondo ciclo, tutte le copie della sequenza target sono state replicate una seconda volta. Il ciclo termico della PCR viene ripetuto molte volte (di solito da 20 a 40 volte), producendo così una replicazione esponenziale della sequenza target. Infatti, ad ogni ciclo successivo il DNA sintetizzato è raddoppiato (Fig. 34).

L’efficienza della PCR dipende dalla capacità di amplificare fedelmente il DNA target, e solo il DNA target. Se la sequenza del DNA target venisse amplificata inserendo nucleotidi sbagliati, allora la PCR potrebbe generare false “mutazioni”, che in realtà non sono presenti nei campioni. Se i primer si agganciassero non solo alla sequenza target, ma anche ad altre sequenze presenti nel DNA dei campioni, allora la PCR amplificherebbe sequenze “aspecifiche” che renderebbero problematica o anche impossibile l’analisi e l’interpretazione dei risultati. L’efficienza e

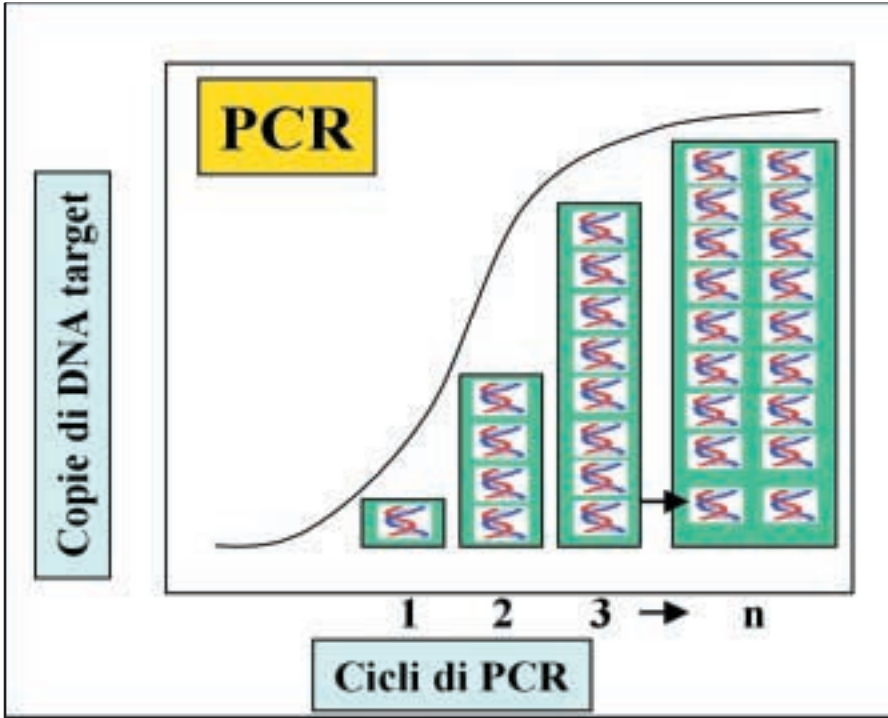


Figura 34 - Amplificazione esponenziale del DNA target nel corso di una PCR.

la specificità della PCR si possono controllare e migliorare ottimizzando caso per caso le condizioni sperimentali. E' molto importante utilizzare Taq polimerasi di ottima qualità. Queste polimerasi sono stabili anche alle alte temperature, quindi non sono degradate quando vengono sottoposte a ripetuti cicli di denaturazione. E' molto importante disporre di primer estremamente specifici, che siano cioè strettamente complementari alle sequenze che fiancheggiano le sequenze target. L'utilizzo di primer specifici consente di evitare l'amplificazione di DNA contaminante. Uno dei fondamentali vantaggi della PCR è che essa consente di amplificare piccolissime quantità di DNA target. Tuttavia, la PCR in presenza di piccole quantità di DNA target è esposta al rischio di contaminazione con DNA estraneo al campione. La contaminazione può avvenire sui campioni biologici prima che il DNA venga estratto in laboratorio, oppure può avvenire in laboratorio nel corso delle operazioni di estrazione o di amplificazione del DNA. Occorre attrezzare il laboratorio ed applicare stringenti controlli di qualità per evitare le contaminazioni,

o per individuarle qualora esse avvengano. L'efficacia della PCR può essere limitata dalla presenza di inibitori della Taq che sono presenti nei campioni. E' possibile predisporre PCR che amplificano contemporaneamente più di un locus (PCR multiplex).

RAPD (Random Amplified Polymorphic DNA)

Sono state recentemente sviluppate altre tecniche di PCR che possono essere utilizzate in genetica forense. La tecnica RAPD (Random Amplified Polymorphic DNA) consiste nell'amplificazione casuale di frammenti di DNA genomico tramite l'uso di primer a sequenza casuale. Le sequenze e la scelta dei primer sono casuali, nel senso che non sono note informazioni sulla presenza di eventuali siti di annealing complementari nel genoma dei campioni analizzati. I primer sono lunghi circa 10 nucleotidi ed hanno sequenze casuali. Il metodo RAPD è una PCR realizzata usando un solo primer ed una bassa temperatura di annealing. Infatti, i primer RAPD sono poco specifici e non potrebbero agganciarsi alle sequenze complementari se non a basse temperature di annealing. Quando due primer RAPD identificano due sequenze complementari entro circa 2.000 nucleotidi di distanza, allora la PCR amplifica frammenti di DNA a filamento doppio, corrispondenti alla regione di DNA inclusa fra i siti di annealing dei due primer. I frammenti di DNA amplificati vengono poi separati in gel di agarosio al 4%, ed evidenziati tramite colorazione con EtBr, oppure in gel di acrilamide al 3 - 5% e colorati con nitrato d'argento ("silver staining"). Il metodo RAPD di solito funziona meglio in DNA vegetali che negli animali. Ogni singolo primer RAPD può amplificare da 3 a 10 frammenti utilizzabili per la diagnosi. Ripetute PCR con diversi primer RAPD possono generare un gran numero di frammenti, che spesso sono molto variabili e polimorfici. Tuttavia questo metodo presenta diverse limitazioni. Di solito i genotipi RAPD non corrispondono ai DNA fingerprinting, perchè i singoli individui non sono distinguibili. Il numero e la qualità dei frammenti amplificati varia a seconda della quantità e qualità del DNA del campione. Il metodo RAPD è molto sensibile a piccole variazioni di una serie di fattori sperimentali, che includono i metodi di estrazione del DNA, le procedure di PCR e di elettroforesi. La tecnica RAPD è quindi poco ripetibile e l'interpretazione dei risultati può richiedere scelte soggettive. La tecnica amplifica sia il DNA del campione che il DNA contaminante. Da un punto di vista genetico, i frammenti RAPD non sono, per definizione, riconducibili a specifici loci, quindi è impossibile o difficile identificare eventuali alleli. Un'ulteriore

complicazione deriva dalle relazioni di dominanza fra frammenti RAPD: un frammento che amplifica su un cromosoma è dominante sul frammento che non amplifica e quindi l'eterozigote non è identificabile. La tecnica RAPD può essere usata per il sessaggio molecolare.

AFLP (Amplified Fragment Length Polymorphism)

Nel metodo AFLP (Amplified Fragment Length Polymorphism) il DNA viene digerito con due enzimi di restrizione, usualmente EcoRI e MseI. I frammenti di restrizione sono poi legati a degli oligonucleotidi che fungono da siti di annealing per i primer (adattatori). Uno specifico adattatore è legato all'estremità 5' ed un'altro all'estremità 3' dei frammenti, tramite l'enzima ligasi. I frammenti con gli adattatori vengono riamplicati usando coppie di primer complementari agli adattatori, a cui sono aggiunti, all'estremità 3', da uno a tre nucleotidi scelti a caso (nucleotidi selettivi). Le combinazioni di primer che si possono usare per le PCR selettive sono moltissime, e devono essere sperimentate ed ottimizzate al fine di ottenere l'amplificazione ripetibile di circa 50 - 100 frammenti. Questi frammenti sono separabili in elettroforesi su gel di sequenza progettati per separare frammenti lunghi al massimo circa 400 - 500 nucleotidi. I frammenti AFLP possono essere visualizzati tramite l'uso di primer marcati radioattivamente (nel caso di elettroforesi manuale), oppure con primer fluorescenti (nel caso di elettroforesi automatiche). Attualmente l'applicazione del metodo AFLP in genetica forense è limitato a casi di identificazione di specie e di ibridi in DNA estratto da campioni di tessuti o di prodotti animali di origine ignota. La tecnica AFLP può essere usata per il sessaggio molecolare.

Sequenziamento del DNA

I metodi oggi in uso per il sequenziamento del DNA vennero sviluppati da Maxam e Gilbert, e da Sanger e collaboratori, nel 1977. Questi metodi usano due approcci diversi per determinare le sequenze di DNA. Oggi le sequenze di DNA vengono fatte quasi esclusivamente utilizzando sequenziatori automatici che si basano sul metodo di Sanger. Il frammento di DNA che deve essere sequenziato viene denaturato. Un primer oligonucleotidico complementare ad uno dei due filamenti di DNA da sequenziare viene utilizzato per avviare la duplicazione del DNA. Nella reazione di sintesi vengono aggiunti i quattro deossinucleotidi (dATP, dGTP, dCTP, dTTP), e porzioni prestabilitate di quattro dideossinu-

cleotidi (ddATP, ddGTP, ddCTP, ddTTP). I dideossinucleotidi sono analoghi modificati dei deossinucleotidi. Quando i dideossinucleotidi sono incorporati nel DNA ne interrompono l'estensione, generando così frammenti che terminano con uno dei quattro dideossinucleotidi. I frammenti vengono poi separati tramite elettroforesi in gel di acrilamide o in sequenziatori automatici a capillari (Fig. 35).

Le reazioni di sequenziamento sono reazioni di PCR effettuate usando un solo primer, cioè il primer di sequenza, ed aggiungendo alla reazione quantità prestabilite di dideossinucleotidi. La scelta del primer di sequenza determina la sequenza che si ottiene, la quale corrisponde ai circa 300 - 800 nucleotidi che sono in posizione 3' rispetto al primer (ricordiamo che la duplicazione del DNA e quindi l'estensione del primer avviene in direzione 5' -> 3'). Poichè nella PCR di sequenziamento si usa un solo primer, l'amplificazione del prodotto sarà lineare e non logaritmica, come invece avviene nelle reazioni di PCR di amplificazione. Il numero di cicli di PCR di sequenziamento dipende dalla quantità di DNA utilizzato, ma di solito non presenta problemi critici. Prima di effettuare l'elettroforesi in un sequenziatore automatico, il prodotto della PCR di sequenziamento deve essere purificato per rimuovere i primer o i terminatori non utilizzati. Una semplice precipitazione del DNA tramite etanolo è di solito sufficiente per ottenere una buona purificazione. Le sequenze di DNA possono essere conservate liofilizzate o congelate a - 20°C per alcuni mesi. Prima dell'elettroforesi, i campioni sono risospesi nel tampone di sequenziamento, che contiene anche formammide, sostanza denaturante. La soluzione di DNA viene denaturata a caldo ed immediatamente caricata nel sequenziatore.

I sequenziatori automatici utilizzano molecole chimiche fluorescenti che vengono legate al DNA e consentono di individuare i frammenti. I sequenziatori automatici sono utilizzati non solo per sequenziare il DNA, ma anche per evidenziare ed analizzare microsatelliti, frammenti generati tramite tecniche RAPD e AFLP; loci VNTR (minisatelliti) e polimorfismi di conformazione del DNA a filamento singolo (Single-Strand Conformational Polymorphism - SSCP). I sequenziatori automatici sono dotati di serie di capillari multipli (di solito 16 o 96), e sono perfettamente adattati per caricare automaticamente le reazioni da analizzare direttamente dalle piastre microtiter. I sequenziatori a capillari non usano marcatori radioattivi ed altre sostanze nocive tipo acrilamide, quindi consentono di eliminare i rischi sanitari sul posto di lavoro. I sistemi di marcatura fluorescenti usano molecole chiamate fluorofori che sono più sensibili dei sistemi di marcatura che usano radioisotopi e sono estremamente più sensibili dei sistemi di

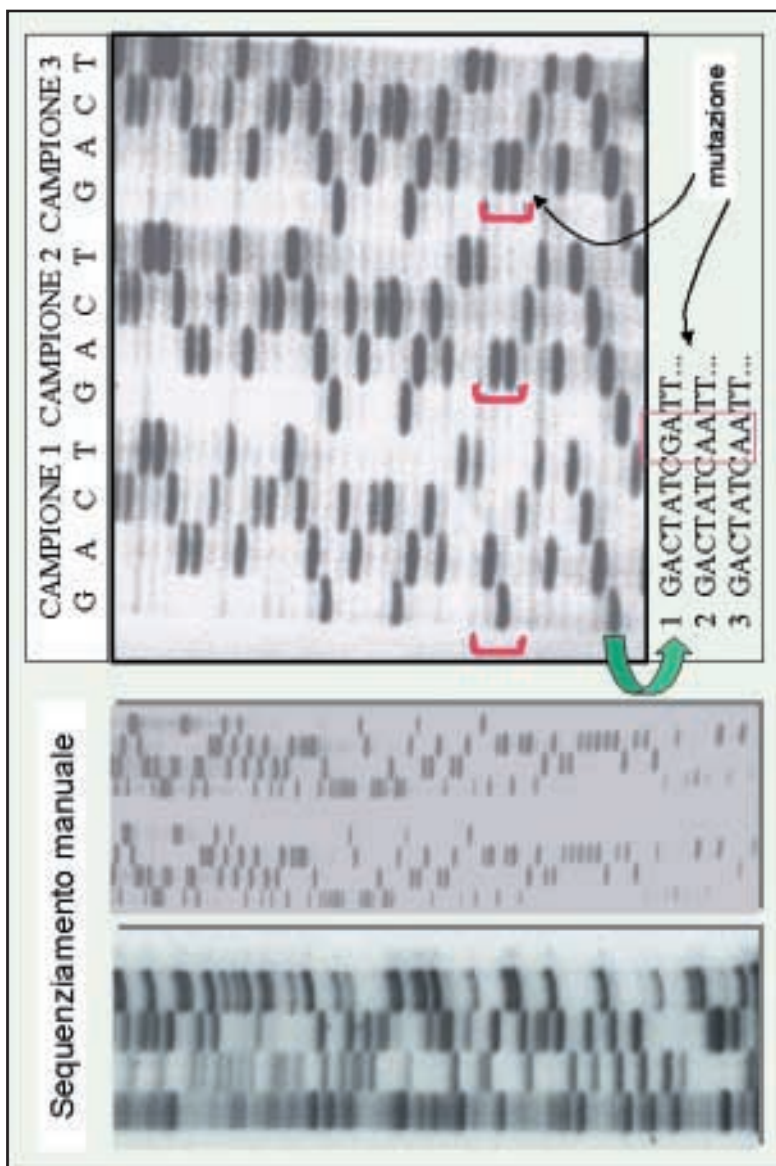


Figura 35 - Sequenziamento manuale del DNA. I prodotti delle quattro reazioni di sequenza, corrispondenti ai quattro nucleotidi del DNA, sono marcati radioattivamente e separati in elettroforesi su gel di acrilamide. L'autoradiografia rivela le sequenze (immagini a sinistra). Comparando visualmente le autoradiografie ottenute dal sequenziamento di campioni diversi, è possibile ricostruire la sequenza del segmento di DNA analizzato, ed è possibile individuare le mutazioni. Nell'immagine a destra si presentano porzioni di sequenze di tre campioni. Nel campione a sinistra si identifica una mutazione da A a G.

I sequenziatori di prima generazione richiedono la preparazione di un gel di acrilamide (di solito un gel denaturante concentrato al 6%). Condizione essenziale per una buona riuscita dell'elettroforesi di sequenziamento è che il gel sia sottile. Pertanto si costruiscono gel di 0.2 - 0.4 millimetri di spessore. E' anche estremamente importante che le lastre di vetro utilizzate per preparare lo stampo per il gel siano assolutamente pulite, e che tutti i prodotti, incluso l'acqua, utilizzati per le soluzioni per il gel siano della massima purezza possibile. Le impurità presenti nei vetri e nel gel possono produrre colori di fondo, che possono mascherare i segnali luminosi dei fluorofori e rendere impossibile la lettura delle sequenze. I sequenziatori di seconda generazione, a capillari, non richiedono la preparazione del gel. Infatti, il sequenziatore inietta automaticamente il polimero nei capillari. In questo modo vengono eliminati tutti i problemi di manualità e risolti tutti i problemi di impurità. I capillari subiscono un processo di usura, che riduce l'accuratezza di lettura dell'elettroforesi. Pertanto ogni set di capillari deve essere sostituito dopo circa 100 - 200 elettroforesi. Le condizioni di elettroforesi sono programmate tramite il software del computer che attiva e controlla tutte le operazioni che sono compiute dal sequenziatore automatico. Quando l'elettroforesi dei campioni è completata, il sequenziatore crea dei files registrati nel computer che contengono tutte le informazioni necessarie per determinare accuratamente le sequenze.

Nel corso di ogni elettroforesi il computer ricostruisce uno o più file d'immagine, che consentono di seguire in tempo reale l'andamento delle analisi. Al termine dell'elettroforesi il file d'immagine viene permanentemente salvato (Fig. 37). I risultati delle analisi di sequenza sono salvati in forma di "elettroferogrammi" (Fig. 36). Quando un fluoroforo è eccitato dal laser produce un'emissione luminosa che viene registrata come picco. L'altezza del picco indica l'intensità dell'emissione ed il colore indica il colore del fluoroforo. Poichè ogni colore è univocamente associato ad una specifica reazione di terminazione, la sequenza dei picchi colorati dell'elettroferogramma corrisponde esattamente alla sequenza del DNA. Il file con l'elettroferogramma contiene la sequenze del DNA scritta con le quattro lettere che corrispondono ai quattro nucleotidi. Le posizioni dei nucleotidi dall'inizio della sequenza sono numerate di 10 in 10. Nel caso in cui un picco particolare non possa essere determinato univocamente, il software segnala la presenza di un'ambiguità (sequenza non determinata = N). La sequenza fornita dall'elettroferogramma può poi essere manipolata tramite l'uso di un apposito software che consente di assegnare manualmente eventuali nucleotidi ambigui. Si possono elimi-

nare porzioni della sequenza che non sono state determinate accuratamente, come di solito accade nella parte iniziale e nella parte finale di ogni elettroferogramma. Le sequenze corrette sono poi trascritte in file di testo. I file di testo vengono allineati con altre sequenze di riferimento, cioè inclusi nei data base ed analizzati.

I prodotti di PCR possono essere sequenziati direttamente. Per ottenere buone sequenze è importante che il DNA sia purificato e non contenga contaminanti che possono inibire la PCR (ad esempio residui di prodotti organici come fenolo e cloroformio, che possono inattivare la Taq polymerase). I prodotti di PCR devono essere specifici e non devono contenere frammenti aspecifici che possono produrre sequenze multiple ed illeggibili. Se i prodotti di PCR contengono contaminanti oppure frammenti aspecifici devono essere purificati tramite elettroforesi in mini-

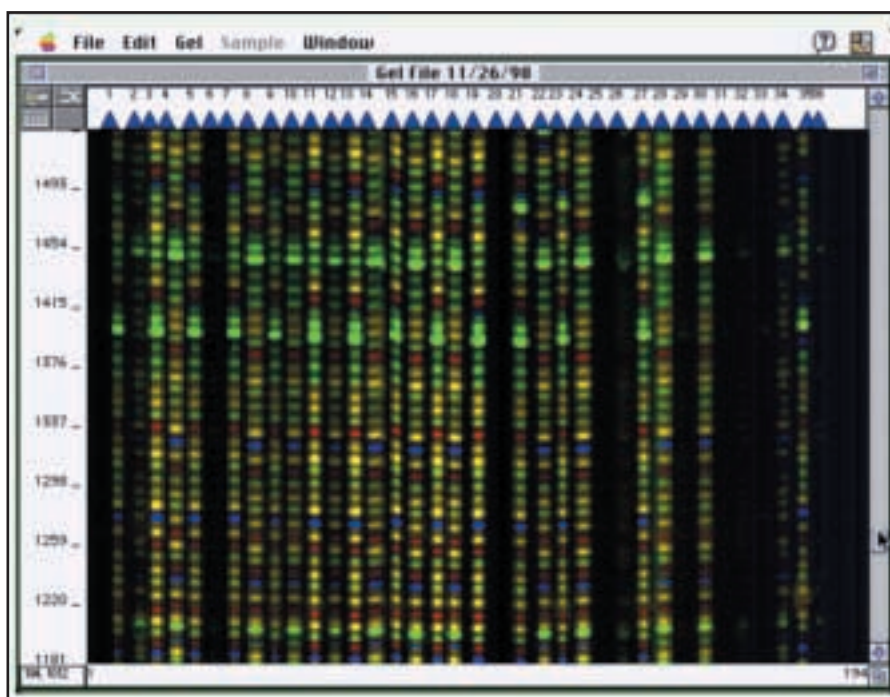


Figura 37 - Immagine di un'elettroforesi di sequenze nucleotidiche in un sequenziatore automatico. Il sequenziamento viene effettuato tramite l'uso di dideossinucleotidi, nucleotidi modificati che terminano la reazione di estensione del primer ogni volta che sono incorporati. I prodotti delle reazioni di sequenziamento sono marcati con quattro colori diversi, corrispondenti ai quattro nucleotidi del DNA. La sequenza dei quattro colori corrisponde, pertanto, alla sequenza nucleotidica.

gel di agarosio. Il frammento specifico viene recuperato, tagliando la porzione di gel che lo contiene. Il frammento viene poi separato dall'agarosio tramite elettro-eluzione, oppure tramite digestione dall'agarosio con l'enzima agarasi. Prima del sequenziamento i prodotti di PCR devono essere purificati, cioè occorre eliminare i nucleotidi ed i primer di PCR che non sono stati utilizzati e che potrebbero interferire con le PCR di sequenziamento. Queste purificazioni sono fatte utilizzando sistemi di filtrazione, con provette dotate di membrane filtranti, oppure con colonnine a resine filtranti, le quali trattengono il DNA a doppia elica prodotto dalla PCR ed eliminano i nucleotidi ed i primer.

Può essere necessario clonare i frammenti amplificati che contengono due alleli (frammenti eterozigoti) e sequenziare i cloni per determinare le differenze fra i due alleli. La presenza di dimeri fra i primer di PCR può produrre artefatti di sequenziamento, soprattutto se uno dei primer di PCR viene anche usato come primer di sequenza. Per eliminare o ridurre questi artefatti è possibile usare tecniche di PCR "hot start", cioè PCR che iniziano ad alta temperatura. L'alta temperatura impedisce la formazione di dimeri e rende più specifica l'interazione fra il primer ed il DNA target. La PCR hot start si realizza utilizzando Taq polimerasi modificate che sono attivate solo ad alta temperatura, quando cioè le interazioni deboli fra dimeri e le interazioni aspecifiche non sono più possibili.

Struttura e sequenziamento del DNA mitocondriale

Il genoma dei mitocondri, cioè il DNA mitocondriale (mtDNA), è una molecola semplice, aploide, che apparentemente non ricombina (ma in alcuni casi può ricombinare). Il mtDNA è una molecola circolare a filamento doppio, generalmente composta da alcune migliaia di nucleotidi. È ereditata, tranne pochi casi, soltanto per via materna. Ogni sequenza di mtDNA, definita un "aplotipo", viene trasmessa intatta nel corso delle generazioni, e quindi può essere utilizzata per ricostruire le genealogie nelle popolazioni e per identificare individui che appartengono a differenti popolazioni, sottospecie e specie. I geni mitocondriali evolvono più rapidamente, in media, dei geni nucleari, e quindi accumulano rapidamente differenze genetiche fra popolazioni di organismi della stessa specie. L'analisi del mtDNA è molto utile per identificare popolazioni che evolvono indipendentemente, gruppi di popolazioni tassonomicamente distinte, casi di ibridazione e di flusso genico.

L'organizzazione del mtDNA è abbastanza stabile. Tutti i DNA mitocondriali dei metazoi contengono geni che codificano per alcune proteine

enzimatiche della catena della respirazione cellulare, geni che codificano per due RNA ribosomali, e circa 20 geni che codificano per RNA di trasferimento (Fig. 11). Inoltre, i genomi mitocondriali possiedono almeno una regione di controllo, che non codifica per proteine o per RNA, ma che ha il ruolo di controllare la replicazione e la trascrizione dell'intero genoma mitocondriale. L'ordine dei geni nel mtDNA è abbastanza conservato, ma esistono numerosi riarrangiamenti. Il DNA mitocondriale evolve 5 - 10 volte più rapidamente dei geni nucleari. In particolare la regione di controllo evolve molto più velocemente delle sequenze nucleari. La regione di controllo contiene alcune brevi sequenze ipervariabili che evolvono molto rapidamente e che sono estremamente utili in genetica delle popolazioni ed in genetica forense. Le sequenze mitocondriali sono facilmente amplificabili utilizzando la PCR e primer specifici. Normalmente, le sequenze mitocondriali vengono analizzate tramite sequenziamento nucleotidico.

Amplificazione ed analisi dei microsatelliti

Le sequenze ripetute dei microsatelliti sono fiancheggiate da sequenze a copia singola. E' così possibile disegnare primer per PCR che consentono di amplificare selettivamente i loci microsatellite. L'analisi dei genotipi si effettua identificando il peso molecolare degli alleli presenti ad ogni locus tramite gel elettroforesi. Come nel caso dei loci VNTR, i genotipi individuali sono determinati analizzando separatamente un certo numero di microsatelliti (di solito 4 - 6) e cumulando i dati, cioè formando genotipi multilocus che corrispondono ai DNA fingerprinting dei sistemi MLP e SLP. I microsatelliti hanno il vantaggio di essere geneticamente ben identificati, di avere un massimo di due alleli per individuo, ma molti alleli nelle popolazioni. I microsatelliti hanno un'ulteriore vantaggio sui loci VNTR: essi sono amplificabili tramite PCR e quindi sono tipizzabili in quasi ogni tipo di campione biologico, indipendentemente dalla concentrazione del DNA e dal suo stato di degradazione.

I microsatelliti presenti nel genoma di una specie sono individuati tramite tecniche di clonaggio, che consentono di isolare quelle sequenze di DNA che includono sia il microsatellite che le sequenze fiancheggianti. Le due sequenze fiancheggianti e la struttura del microsatellite sono identificati tramite sequenziamento nucleotidico. L'analisi delle sequenze fiancheggianti consente di disegnare i primer per PCR. La variabilità espressa da ogni nuovo locus microsatellite deve essere caratterizzata in un campione di individui prelevati dalle popolazioni di riferimento. Infatti, non

tutti i loci microsatellite sono polimorfici. L'analisi dei microsatelliti viene fatta separando gli alleli tramite elettroforesi in gel denaturanti (gel di sequenziamento), che consentono di separare chiaramente i due alleli presenti ai loci eterozigoti (Fig. 15). L'elettroforesi deve essere eseguita con tecniche estremamente precise, poiché la differenza di peso molecolare fra gli alleli dipende frequentemente da due (nei microsatelliti dinucleotidici), tre, quattro o sei nucleotidi. Non tutti gli alleli microsatellite sono costituiti da un repeat perfetto e, talvolta, le differenze fra due alleli sono dovute ad un singolo nucleotide. Quindi, il sistema elettroforetico deve consentire di separare ed identificare frammenti (alleli) che differiscono per un solo nucleotide, esattamente come nei gel di sequenziamento. In genetica forense è consigliabile analizzare i microsatelliti utilizzando sequenziatori automatici su gel di acrilammide o in sistemi a capillari.

Analisi dei microsatelliti nei sequenziatori automatici

Nei sequenziatori automatici è possibile analizzare contemporaneamente diversi loci microsatellite in uno stesso capillare. L'analisi di loci multipli si può realizzare sia tramite PCR multiplex sia tramite elettroforesi di PCR singole cumulate (elettroforesi multiplex). Nei sistemi multiplex (sia di PCR che di elettroforesi) occorre scegliere loci microsatellite che producano segnali (elettroferogrammi) puliti e netti, con il minor numero possibile di segnali aspecifici. Nell'analisi automatica dei microsatelliti uno dei due primer di PCR viene marcato con un fluoroforo. Nei sistemi multiplex occorre marcare i primer ai diversi loci con colori diversi. Attualmente si usano tre colori (giallo, verde e blu) per la marcatura dei primer, mentre il quarto colore (rosso) viene utilizzato per marcare il peso molecolare standard. La sintesi e le marcature dei primer vengono effettuate da laboratori commerciali. Nei sistemi multiplex si combinano microsatelliti i cui alleli hanno pesi molecolari differenti. Così i prodotti di PCR sono separati in zone diverse del gel o del capillare, e l'identificazione degli alleli è facilitata dalla lettura di segnali dai colori diversi che non si sovrappongono (Fig. 38). Si possono marcare con lo stesso colore microsatelliti che hanno alleli con pesi molecolari che differiscono di almeno 50 nucleotidi. Il software di analisi dei risultati dell'elettroforesi genera un file d'immagine (Fig. 38) ed un elettroferogramma (Fig. 39). Gli elettroferogrammi delle analisi di microsatelliti sono evidentemente più semplici degli elettroferogrammi di sequenza. I pesi molecolari degli alleli sono determinati con precisione tramite l'uso degli standard interni. Ogni allele può essere composto da una sin-

gola banda (che appare come un singolo picco nell'elettroferogramma), oppure dalla banda principale più una serie di bande secondarie che rappresentano prodotti aspecifici di amplificazione. Occorre quindi correggere manualmente o automaticamente il risultato dell'analisi, individuando il segnale prodotto dalla banda principale ed assegnando il rispettivo peso molecolare. Esistono software che consentono di assegnare automaticamente il peso molecolare corretto. Occorre definire il range di variazione del peso molecolare e del picco principale dell'elettrofero-

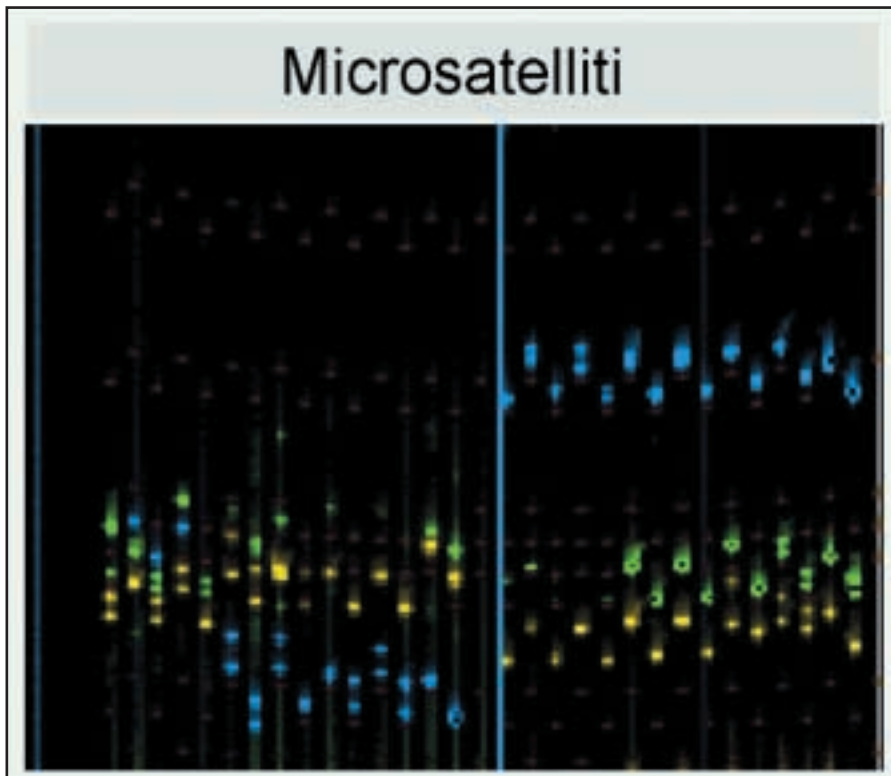


Figura 38 - Immagine di elettroforesi di loci microsatelliti, ottenuta tramite analisi al sequenziatore automatico. Si sono analizzati contemporaneamente quattro diversi microsatelliti: un locus è marcato con il colore giallo; un locus con il colore verde; due loci (con alleli di diverso peso molecolare) sono marcati con colore blu). Gli individui omozigoti presentano un solo frammento colorato ad ogni locus; gli individui eterozigoti presentano due frammenti colorati ad ogni locus. L'immagine del gel è approssimativa. L'analisi dei risultati viene poi successivamente effettuata utilizzando un'apposito software che consente di determinare esattamente i pesi molecolari dei singoli alleli, e quindi di identificare precisamente i genotipi di ogni individuo.

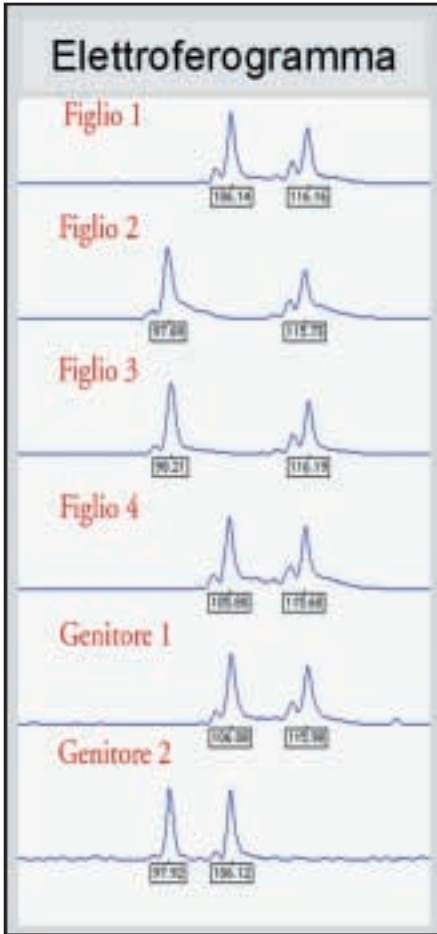


Figura 39 - Elettroferogramma degli alleli ad un locus microsatellite, ottenuto tramite analisi automatica. I picchi dei tracciati indicano la presenza degli alleli, i cui pesi molecolari sono stimati accuratamente. Gli alleli presenti nei due genitori (entrambi eterozigoti) sono presenti, in combinazioni diverse, anche nei figli (che sono tutti eterozigoti).

gramma ed il colore del locus. Il programma quindi usa un algoritmo di filtro delle informazioni che consente di ignorare i segnali secondari ed assegnare il peso molecolare corretto al segnale principale dell'allele. Il risultato finale può essere visualizzato come elettroferogramma corretto, oppure i dati, cioè i valori dei pesi molecolari assegnati a ciascun allele, possono essere esportati in fogli di data-base tipo Microsoft Excel, oppure nei fogli di input dei vari software di elaborazione dei dati. Le stime dei pesi molecolari sono espresse con numeri frazionari (per esempio: 110.53). Ovviamente, poiché le differenze di peso molecolare fra gli alleli sono determinate dal numero di ripetizioni del repeat, i pesi molecolari devono essere non frazionari e devono variare in step di due (o multipli di due) o quattro (o multipli di quattro) nei microsatellidi composti da di- o tetra-nucleotidi, rispettivamente. L'assegnazione di pesi molecolari frazionari dipende dalla estrema precisione delle assegnazioni rispetto allo standard ed alle concomitanti variazioni della velocità di migrazione delle molecole nel

gel. L'entità di queste variazioni sperimentali fa sì che lo stesso allele possa migrare con differenze corrispondenti a ± 1 nucleotide in corse ripetute nello stesso gel. Occorre pertanto aggiustare i pesi molecolari individuati dagli elettroferogrammi, stabilendo un range di variazione ed assegnando

gli alleli incorporandoli in classi di variazione dei pesi molecolari (“bins”). In questo modo sono assegnati allo stesso allele elettroferogrammi che presentano un intervallo di variazione determinato sperimentalmente. Questo sistema è del tutto analogo al sistema di “binning” usato nella determinazione dei pesi molecolari dei frammenti di DNA fingerprinting MLP e VNTR.

Cromosomi sessuali e determinazione genetica del sesso

Le cellule dei tessuti contengono un numero stabile di coppie di cromosomi che hanno forma definita e riconoscibile al microscopio, numero che è caratteristico della specie. Il cariotipo di ogni individuo comprende un certo numero di coppie di cromosomi fra di loro simili (autosomi) ed una sola coppia di cromosomi di forma chiaramente diversa (eterocromosomi). Gli eterocromosomi sono anche detti cromosomi sessuali, perché contengono sequenze di DNA che determinano il sesso dell'individuo. I cromosomi sessuali dei mammiferi sono chiamati X e Y, quelli degli uccelli W e Z. Nei mammiferi i maschi possiedono un cromosoma X ed un cromosoma Y (cariotipo eterogametico XY), mentre le femmine hanno due cromosomi X (cariotipo omogametico XX). Al contrario, negli uccelli i maschi hanno cariotipo ZZ (omogametico) e le femmine ZW (eterogametico). Le cellule-uovo dei mammiferi contengono solo cromosomi X, mentre metà degli spermatozoi contengono cromosomi X e l'altra metà contengono cromosomi Y. Quindi, il sesso dei mammiferi è determinato per via paterna, ed il cromosoma Y è ereditato solo per via paterna. Al contrario, il DNA mitocondriale è ereditato solo per via materna. La determinazione cromosomica del sesso negli uccelli funziona esattamente al contrario di come funziona nei mammiferi.

La presenza di sequenze di DNA uniche, presenti solo nei cromosomi sessuali, consente di effettuare la determinazione molecolare del sesso, e quindi consente di assegnare un sesso a campioni biologici di origine ignota. In molti rettili (in tutte le specie di coccodrilli, nella maggior parte delle tartarughe ed in alcune specie di lucertole) la determinazione del sesso è controllata dall'ambiente, principalmente dalla temperatura ambientale, e quindi non esistono sequenze di DNA che consentono l'identificazione molecolare del sesso. Alcune specie di pesci non hanno sessi distinti: lo stesso individuo può agire da maschio o da femmina. Queste specie non hanno marcatori genetici per il sesso. Tuttavia, molte specie di pesci, rettili ed anfibi potrebbero in futuro essere sessabili per via molecolare, anche se attualmente i marcatori genetici non sono stati

individuati. Nei mammiferi e negli uccelli esistono sempre cromosomi sessuali morfologicamente distinti, quindi è possibile individuare geni e sequenze di DNA non codificanti legate al sesso. Nelle specie a chiara determinazione genetica del sesso gli eterocromosomi sono, di solito, morfologicamente distinti ed hanno sequenze di DNA in parte diverse. Ovviamente le sequenze di DNA associate al sesso omogametico (determinato dai cromosomi X e Z nei mammiferi e negli uccelli, rispettivamente) saranno presenti anche nel cariotipo del sesso eterogametico, mentre le sequenze associate al cromosoma Y e W dei mammiferi ed uccelli saranno presenti unicamente nei maschi dei mammiferi e nelle femmine degli uccelli, e quindi possono essere usate come marcatori molecolari. Le sequenze legate ai cromosomi Y o W possono essere sequenze di geni codificanti che agiscono direttamente nella determinazione del sesso nel corso dello sviluppo embrionale e dell'accrescimento, sequenze anonime non codificanti, oppure sequenze di DNA ripetitivo, cioè di micro- o minisatelliti. Le sequenze di geni codificanti spesso sono conservate e quindi possono essere identificate in gruppi di specie filogeneticamente vicine. Sequenze anonime possono essere identificate tramite metodi RAPD o AFLP, comparando i frammenti di amplificazione che si ottengono in individui maschi e femmine, e cercando la presenza di frammenti che sono associati univocamente ad uno dei due sessi. Questi frammenti non devono essere variabili entro sesso, cioè devono sempre essere presenti, ad esempio, in tutti i maschi ed essere sempre assenti in tutte le femmine. Le sequenze di DNA ripetitivo possono essere individuate tramite PCR (per i microsatelliti che mappano sul cromosoma Y o W), oppure tramite Southern blotting (per i minisatelliti). La probabilità che un marcatore sia specifico per il sesso e non sia semplicemente un marcatore polimorfico nella specie, è determinata tramite la formula:

$$p = q^m(1 - q)^f$$

dove: m = numero dei maschi analizzati; f = numero delle femmine analizzate; $q = m/(m + f)$. Per esempio, se qualora vengano correttamente identificati quattro maschi e quattro femmine, la probabilità che il marcatore sia specifico per il sesso è $p = 0.996$.

Le procedure d'analisi del DNA tramite PCR utilizzate per il sessaggio molecolare devono comunque includere sia controlli negativi, cioè PCR eseguite senza DNA del campione, per escludere che sia sessato DNA contaminante, che controlli positivi che assicurano che la PCR funzioni. Per esempio, se il test prevede che la PCR produca l'amplificazione di un frammento di DNA nei maschi, ma nessun frammento nelle femmine, una PCR che non abbia funzionato verrebbe classificata come femmina

anche se il campione è di maschio. Un controllo positivo richiede l'organizzazione di PCR che includono i primer per amplificare un frammento di DNA, di solito una sequenza di mtDNA, che è presente in ogni campione indipendentemente dal sesso.

E' possibile effettuare il sessaggio molecolare tramite il metodo RAPD. Un frammento RAPD che sia presente in tutti i maschi (nei mammiferi) o in tutte le femmine (negli uccelli) potrebbe essere legato ai cromosomi Y e W, e può essere usato come marcatore per il sessaggio molecolare. Ogni primer RAPD di solito produce l'amplificazione di 1 - 10 frammenti di DNA. E' improbabile che un singolo primer consenta di individuare frammenti legati al sesso. Quindi, occorre analizzare i risultati di un certo numero di primer (10 - 40) per evidenziare eventuali marcatori legati al sesso. Poiché il metodo RAPD notoriamente fornisce risultati che risentono della variabilità legata ai campioni usati per l'estrazione del DNA, ai metodi di estrazione del DNA, alla qualità e concentrazione del DNA, alle condizioni di PCR e di elettroforesi, è necessario standardizzare accuratamente i protocolli di laboratorio, ed assicurarsi che i risultati ottenuti siano ripetibili. La diagnosi di un'elettroforesi di frammenti RAPD dovrebbe essere confermata da almeno due persone che hanno esaminato i gel indipendentemente. Il metodo AFLP è più complicato e più costoso del metodo RAPD, ma consente di generare un numero maggiore di frammenti di DNA, quindi offre maggiori possibilità di individuare frammenti legati al sesso. Marcatori sessuali RAPD e AFLP dovrebbero amplificare bene e produrre frammenti che siano individuabili chiaramente al transilluminatore o nell'elettroferogramma. In entrambi i metodi si dovrebbero identificare frammenti di controllo, cioè frammenti RAPD o AFLP non polimorfici, che amplificano sempre con intensità simile ai marcatori sessuali. Nel caso in cui certi campioni non presentino né i frammenti di controllo né i marcatori sessuali, allora le PCR devono essere rifatte. I marcatori sessuali RAPD e AFLP possono essere clonati e sequenziati al fine di individuare sequenze per disegnare primer per PCR. Tramite PCR si possono amplificare selettivamente i frammenti di DNA legati al sesso. Quindi, le diagnosi di sessaggio molecolare possono essere effettuate direttamente senza ricorrere ai metodi RAPD o AFLP.

Geni legati al cromosoma Y nei mammiferi. Recentemente sono stati scoperti nell'uomo dei geni che determinano il sesso maschile. Questi geni mappano sul cromosoma Y, sebbene sul cromosoma X possano essere presenti geni strutturalmente simili. I geni umani SRY (sex-determining region Y), ZFY (zink finger Y) e AMGY (amelogenina Y) sono presenti

anche sul cromosoma Y di altre specie di mammiferi. Questi geni hanno controparti modificate (SRX, ZFX e AMGX) nel cromosoma X umano e di altre specie di mammiferi. Geni legati al cromosoma W negli uccelli. Gli uccelli hanno due geni CHD1 (chromo-helicase-DNA binding), che mappano sui cromosomi W (CHD1-W) e Z (CHD1-Z), rispettivamente. Sono stati identificati anche i geni ATPase W, ed EEO.6. Per tutti questi geni esistono primer di PCR che consentono di effettuare diagnosi di sesso molecolare (Fig. 40).



Figura 40 - Esempio di sesso molecolare effettuato tramite PCR di sequenze di DNA presenti solo sui cromosomi X ed Y. Nelle femmine (F) è presente un sito di restrizione, che è assente nei maschi (M). Le sequenze amplificate sono digerite con l'enzima di restrizione appropriato: le sequenze dei maschi non sono digerite, mentre le sequenze delle femmine sono digerite. Il DNA digerito è separato in elettroforesi su gel di agarosio, che rivela la presenza di uno (M) o due (F) frammenti. c indica i controlli, campioni amplificati in assenza di DNA. L'assenza di amplificazione indica l'assenza di contaminazioni.

Determinazione della similarità fra frammenti di DNA, alleli, genotipi (aplotipi e genotipi multilocus) ed individui

Le procedure di analisi molecolare producono frammenti di DNA, che spesso sono identificabili come alleli di loci anonimi o ben identificati, da cui si ricostruiscono genotipi (per esempio, sequenze, cioè aplotipi di mtDNA, oppure genotipi multilocus), che sono associati agli individui da cui provengono i campioni analizzati. Tutti questi passaggi, che intercorrono fra l'identificazione dei singoli frammenti di DNA e l'individualizzazione dei campioni, presuppongono l'esecuzione di molteplici operazioni di identificazione che stabiliscono il grado di corrispondenza ("match") fra alleli, genotipi ed individui. Nelle diagnosi di genetica forense, il termine "corrispondenza" significa che non si sono osservate differenze fra i campioni sottoposti al test. E' certamente possibile che due campioni siano differenti, ma che i test utilizzati non abbiano rilevato le differenze. Poiché le analisi genetiche esaminano una parte molto piccola del genoma, ulteriori analisi potrebbero rilevare differenze e portare a differenti conclusioni. L'individualizzazione non ha un valore assoluto. Una conclusione di "corrispondenza" (similarità, concordanza genetica) descrive semplicemente il fatto che, nei particolari test che sono stati condotti, non si sono osservate differenze fra due campioni. Ovviamente i test genetici devono effettuare analisi di sequenze di DNA variabili al punto da "assicurare" che due individui diversi siano sempre identificabili. In genetica forense è quindi di fondamentale importanza valutare il potenziale di individualizzazione dei marcatori genetici e delle procedure di analisi utilizzate.

Identificazione dei frammenti di DNA in analisi di DNA fingerprinting con sonde multilocus

L'elettroforesi in gel di agarosio e l'autoradiografia non consentono di separare frammenti di DNA che differiscono per pochi nucleotidi. Inoltre, numerose cause tecniche possono generare variazioni nella migrazione elettroforetica. Nell'analisi dei sistemi multilocus tramite ibridazione di DNA genomico digerito con endonucleasi con una sonda marcata, è difficile risolvere con precisione i genotipi individuali, perchè è difficile identificare esattamente il numero di repeat di cui è composto ogni frammento. In questi sistemi i singoli loci non sono individuabili, gli alleli ad ogni locus possono essere moltissimi e possono differire pochissimo l'uno dall'altro. Per esempio, un'allele che contiene 99 ripetizioni di un'unità ripetuta composta da 20 nucleotidi, può essere indistinguibile da un'allele che

contiene 100 ripetizioni della stessa unità. La stima delle dimensioni, cioè del peso molecolare (MW) dei frammenti può essere fatta visualmente oppure tramite analisi al computer. L'analisi visuale consente di determinare se due DNA fingerprinting sono chiaramente differenti o simili. Se sono chiaramente differenti, non c'è necessità di procedere ad ulteriori analisi e si può concludere che i due campioni non corrispondono. Se i due profili sono simili, allora i singoli frammenti possono essere analizzati al computer per identificare precisamente le corrispondenze. Il computer analizza l'immagine e determina il MW di ogni singolo frammento; la stima del MW sarà associata, come ogni misura, ad un certo grado di imprecisione. Perciò quando due frammenti sono definiti "corrispondenti", questo non significa che sono esattamente identici, ma che sono identici nell'ambito del margine di errore sperimentale associato con la determinazione del MW. IL MW viene determinato comparando il MW di ogni frammento al MW di un campione di riferimento a MW noto e con uno o più standard di peso molecolare. La corrispondenza fra i MW del campione e quelli dello standard di riferimento non sarà mai esatta, ma sarà corrispondente più o meno un certo errore, definito dalla deviazione in percentuale dal MW del frammento di riferimento. La misura dei frammenti è precisa entro una predefinita percentuale di scostamento. Sulla base delle analisi di qualità delle procedure utilizzate nel laboratorio di genetica forense, occorre definire quale sia la percentuale di scostamento accettabile. Per esempio, due frammenti sono dichiarati corrispondenti se hanno lo stesso MW $\pm 2.5\%$ di scostamento dal loro MW medio (Fig. 41). Due frammenti corrispondenti vengono assegnati allo stesso "allele". I DNA fingerprinting MLP di due campioni vengono considerati corrispondenti se tutti i loro frammenti sono assegnati agli stessi alleli.

Stima delle frequenze alleliche tramite binning in sistemi multilocus

Nei sistemi RFLP non è possibile definire con esattezza l'allele, ed è quindi impossibile definire le frequenze alleliche nelle popolazioni di riferimento. Per definire le frequenze alleliche nei sistemi MLP si ricorre al sistema del binning. I bins sono intervalli prestabiliti di variazione del MW di ogni frammento, che definiscono classi (bins) di alleli: tutti i frammenti ("alleli") che sono inclusi entro un determinato intervallo di variazione, sono assegnati ad un singolo bin. Due frammenti di DNA, in due campioni, che migrano all'interno dell'intervallo di variazione prestabilito, sono assegnati allo stesso bin. Tutti i frammenti assegnati allo stesso bin vengono considerati per calcolare la distribuzione di frequenza

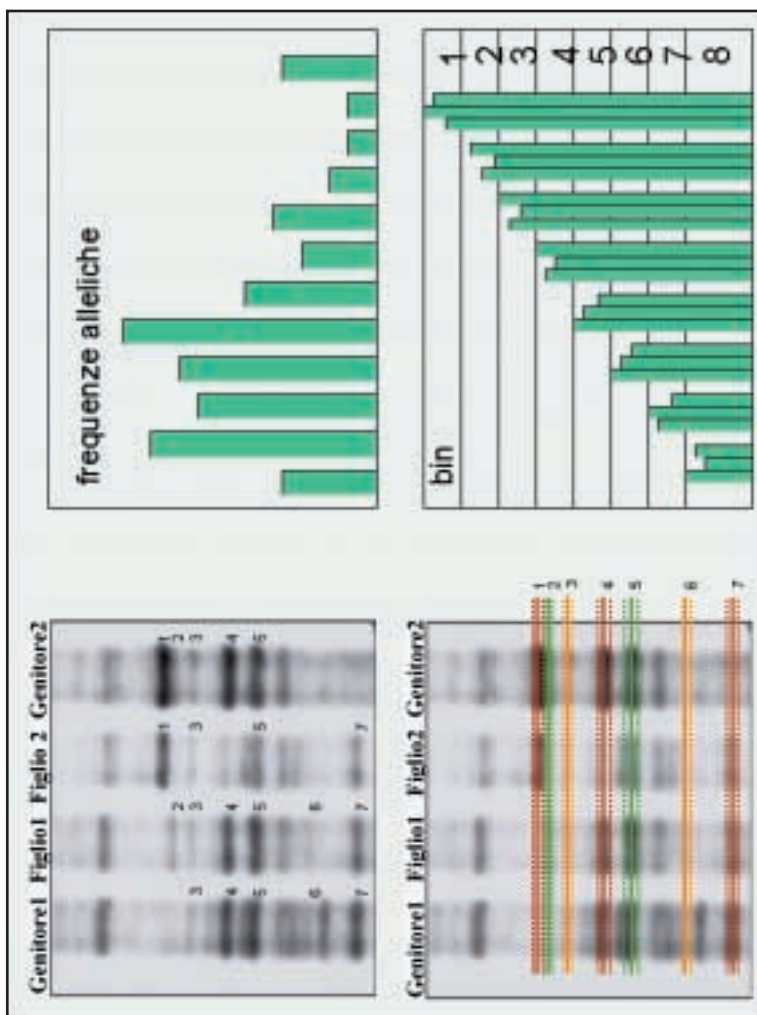


Figura 41 - Analisi di DNA fingerprinting tramite sonde multilocus (MLP). A destra è rappresentata l'autoradiografia di una famiglia di psittaciformi. I due figli presentano frammenti di DNA (alleli) che sono presenti anche in uno o nell'altro dei due genitori putativi. Tuttavia, a causa dei limiti del metodo di analisi (elettroforesi di frammenti RFLP in gel di agarosio), i pesi molecolari di frammenti corrispondenti non sono esattamente identici. L'identificazione degli alleli avviene con il metodo del "binning". Si stabilisce un intervallo di variazione (bin) attorno al peso molecolare medio di ogni allele (figura in basso a sinistra). Tutti i frammenti che rientrano nell'intervallo di variazione vengono identificati come appartenenti allo stesso bin (istogramma in basso a destra). Due frammenti corrispondenti vengono definiti "corrispondenti", non sono esattamente identici, ma sono identici nell'ambito del margine di errore sperimentale associato con la determinazione del peso molecolare. Due frammenti corrispondenti vengono assegnati allo stesso "allele". I DNA fingerprinting di due campioni vengono considerati corrispondenti se tutti i loro frammenti sono assegnati agli stessi alleli.

dei bin nelle popolazioni di riferimento (Fig. 41). Nella stima delle probabilità di identità si considerano le frequenze dei bin a cui appartengono gli “alleli”, non le frequenze degli “alleli”. In genetica forense umana ogni bin deve contenere almeno 5 alleli diversi. In genetica forense di specie CITES quasi sempre mancano adeguati data-base delle popolazioni di riferimento. E' quindi difficile costruire un sistema di binning. Perciò gli alleli vengono identificati sulla base dell'intervallo di variazione dei singoli frammenti, ogni allele così definito corrisponde ad un bin, le frequenze alleliche e le frequenze dei bin nelle popolazioni di riferimento sono equivalenti.

In genetica forense umana i sistemi MLP sono stati rimpiazzati nel 1990 dall'analisi dei minisatelliti con sonde a locus singolo (VNTR-SLP).

Identificazione degli alleli in analisi di DNA fingerprinting con sistemi VNTR

Nei sistemi VNTR sono tipizzati loci singoli, cui è possibile attribuire chiaramente gli alleli che compongono i genotipi individuali (Fig. 42). Le frequenze alleliche possono essere stimate esplicitamente. Tuttavia, anche nei sistemi VNTR l'elettroforesi in gel di agarosio e l'autoradiografia non consentono di separare frammenti di DNA che differiscono per pochi nucleotidi, ed il MW degli alleli non può essere misurato con accuratezza superiore al 2 - 3%. E' possibile descrivere falsi omozigoti: una singola banda ad un locus può derivare da un “vero” omozigote, oppure da un eterozigote con due alleli di MW troppo simile per poter essere distinti. In alcuni casi non è quindi possibile definire con esattezza gli alleli che compongono i genotipi, ed occorre ricorrere al sistema del binning.

Ad iniziare dal 1994 i sistemi SLP sono stati progressivamente sostituiti dall'analisi dei loci microsatellite.

Identificazione degli alleli in analisi di DNA fingerprinting con microsatelliti (STR)

I sistemi STR, amplificati tramite PCR e tipizzati tramite elettroforesi su gel di acrilammide, consentono di identificare il MW dei singoli alleli presenti ai singoli loci con maggiore precisione, soprattutto se l'elettroforesi viene fatta nei sequenziatori automatici. Gli elettroferogrammi (vedi analisi del DNA tramite sequenziatori automatici) consentono di determinare il MW degli alleli STR con grande precisione. I loci microsatellite

di solito hanno un numero di alleli per locus inferiore agli alleli presenti ai loci minisatellite. Inoltre gli alleli STR sono meglio caratterizzabili, sapendo che possono variare di due, quattro o sei nucleotidi, in loci composti da unità ripetute di dinucleotidi, tetranucleotidi o esanucleotidi.

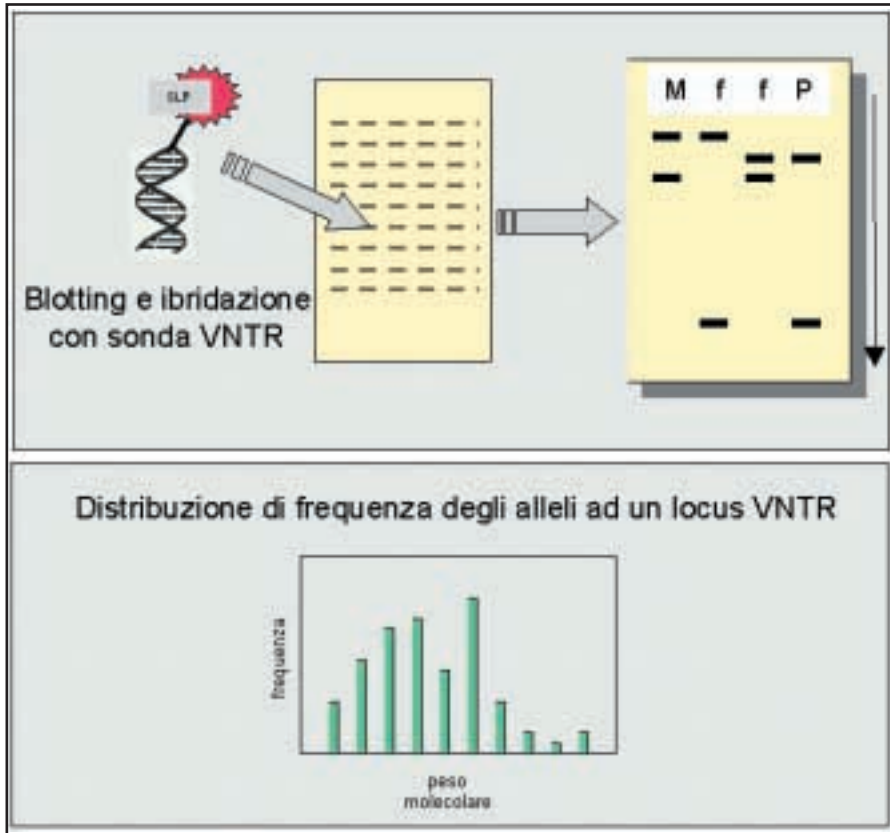


Figura 42 - I sistemi VNTR consentono di tipizzare loci singoli, cui è possibile attribuire chiaramente gli alleli che compongono i genotipi individuali. Ogni sonda VNTR identifica gli alleli presenti ad un singolo locus. Le frequenze alleliche possono essere stimate esplicitamente. Tuttavia, anche nei sistemi VNTR l'elettroforesi in gel di agarosio e l'autoradiografia non consentono di separare frammenti di DNA che differiscono per pochi nucleotidi, ed il peso molecolare degli alleli non può essere misurato con accuratezza superiore al 2 - 3%. Nello schema a destra si presenta un esempio di elettroforesi del DNA di due genitori (M e P) e dei due figli (f). I due genitori sono eterozigoti, ed i figli presentano combinazioni diverse degli alleli presenti nei due genitori. La freccia indica la direzione della migrazione elettroforetica. Gli alleli verso il basso sono gli alleli più leggeri che migrano più velocemente degli alleli più pesanti (in alto). Nella parte inferiore della figura è rappresentata una distribuzione di frequenza degli alleli presenti ad un locus VNTR in una popolazione di riferimento.

Perciò i loci STR sono ben caratterizzabili: la struttura genetica del repeat può essere descritta tramite sequenziamento nucleotidico, gli alleli presenti nelle popolazioni possono essere esattamente caratterizzati descrivendo il numero di repeat di cui sono composti e quindi definendo esattamente la loro lunghezza ed il loro MW; i loci possono essere mappati sui cromosomi, e quindi le loro reciproche relazioni di linkage possono essere precisamente definite. Restano dei problemi: primo fra tutti i rischi di contaminazione, che nei sistemi MLP sono virtualmente inesistenti; poi il problema del dropout allelico.

ANALISI STATISTICA DEI DATI

Distribuzioni di frequenza

I risultati delle analisi del laboratorio di genetica forense (identificazione dei genotipi individuali) sono utilizzati per determinare le probabilità di individualizzazione dei campioni e per costruire le distribuzioni delle frequenze alleliche nelle popolazioni cui i campioni appartengono. Nella valutazione di questi risultati possono sorgere svariati problemi. Per esempio: qual'è l'incertezza associata alle stime delle frequenze alleliche nelle popolazioni di riferimento? In che modo si possono utilizzare le frequenze alleliche per stimare correttamente le frequenze dei genotipi? Le frequenze genotipiche osservate sono una buona stima delle frequenze nella popolazione di riferimento? Qual è la probabilità di osservare un genotipo G in un individuo campionato a caso dalla popolazione di riferimento? La teoria delle probabilità ed i metodi dell'analisi statistica forniscono gli strumenti per effettuare correttamente queste valutazioni. L'analisi statistica fornisce le procedure di calcolo per stimare correttamente i valori dei parametri della popolazione utilizzando i dati ottenuti dai campioni. Un campione dovrebbe essere casuale, rappresentativo, stratificato ed avere le dimensioni appropriate. Tuttavia, raramente, nel corso delle procedure di genetica forense, è possibile lavorare su campioni rappresentativi provenienti da popolazioni di riferimento ben caratterizzate. Spesso si usano i dati disponibili nelle banche dati.

I dati statistici sono sempre ottenuti da campioni di dimensioni limitate estratti da popolazioni che comprendono tutti gli individui esistenti. Un campione casuale è un sottoinsieme di individui scelti casualmente, cioè scelti in modo che ognuno dei componenti la popolazione abbia esattamente la stessa probabilità di essere incluso nel campione. La

dimensione del campione è indicata con n . Il campione è misurato per ottenere i valori dei caratteri X_i prescelti per descriverlo (per esempio, le frequenze alleliche ad una serie di microsattelliti). I valori che il carattere X prende nel campione sono detti osservazioni, e sono indicati con $x_1, x_2, x_3, \dots, x_n$. Le osservazioni ed i dati statistici sono rappresentati in forma grafica o tabulare. Una tabella di frequenze alleliche può essere rappresentata graficamente come distribuzione di frequenza o istogramma (Fig. 43). Le osservazioni hanno una distribuzione di frequenza, che è descrivibile in termini matematici (ad esempio, la distribuzione normale, la distribuzione di Poisson). Ogni distribuzione di frequenza è descritta dalle stime dei “parametri” della popolazione da cui i dati sono stati estratti, e cioè dalle misure delle tendenze centrali della distribuzione (media, mediana e moda) e dalle misure di dispersione delle osservazioni attorno alla media (varianza, deviazione standard, percentili).

Le procedure di analisi statistica consentono di stimare i parametri delle distribuzioni, di valutare la precisione delle stime e di saggiare la significatività delle ipotesi. Le stime dei parametri sono ottenute tramite l’analisi dei campioni. La statistica opera comparando i valori osservati con i valori attesi ottenuti dalle distribuzioni di frequenza. Una distribuzione di frequenza assegna una probabilità ad ogni possibile valore delle variabili. Per esempio, nel lancio di una moneta si assume che la probabilità di realizzazione dell’evento “testa” sia uguale alla probabilità dell’evento “croce” (50%). In genetica delle popolazioni ed in genetica forense, si assume che gli alleli ed i genotipi osservati siano stati estratti da una popolazione con **distribuzione binomiale** (di Bernoulli), che descrive la distribuzione di frequenza di una serie di eventi indipendenti, ognuno dei quali ha due possibili risultati: a_1 (con probabilità p) oppure a_2 (con probabilità $q = 1 - p$). Se $p = q = 0.50$, possiamo calcolare le probabilità di ottenere gli eventi: $a_1a_1 = p \times p = 0.25$; $a_1a_2 = a_2a_1 = 2 \times p \times q = 0.50$; e $a_2a_2 = q \times q = 0.25$, che sono corrispondenti alle proporzioni dei genotipi attesi sulla base delle leggi di Mendel. Se $p \neq q$, la distribuzione binomiale descrive le proporzioni genotipiche attese in una popolazione in equilibrio di Hardy-Weinberg. La distribuzione binomiale è definita dalla funzione di probabilità (pdf):

$$\Pr(x|n, p) = [n! / x!(n-x)!] p^x (1-p)^{(n-x)}$$

La probabilità (Pr) che un’evento (che ha probabilità p) si verifichi x volte in un campione di dimensione n , è uguale alla pdf, con:

$n!$, $x!$, e $(n-x)!$ = n , x , e $(n-x)$ fattoriali (per esempio, se $n = 3$, $3! = 3 \times 2 \times 1 = 6$).

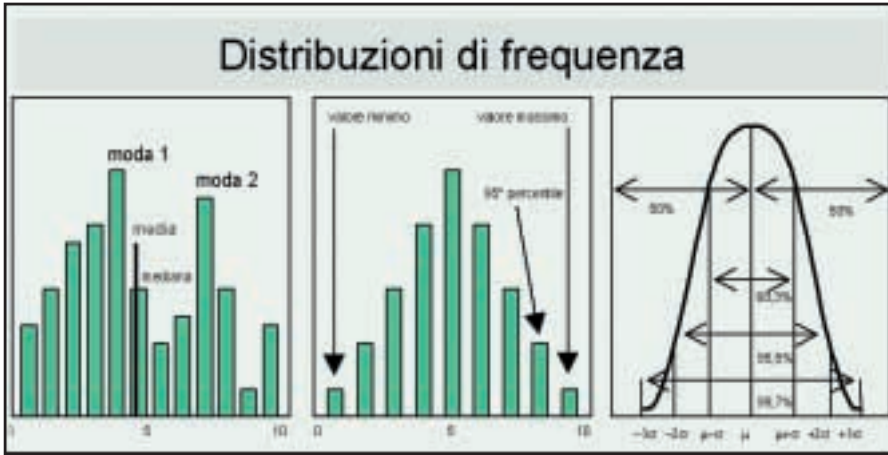


Figura 43 - Istogramma (a sinistra), distribuzione di frequenza (al centro) e distribuzione normale standardizzata (a destra). L'istogramma presenta i dati raggruppati in dodici classi di frequenza. Ci sono due classi di maggior frequenza (moda), quindi la distribuzione è bimodale. Sono indicati i valori della media aritmetica, e della mediana (il valore centrale delle osservazioni nel campione: 50% dei valori sono inferiori e 50% sono superiori alla mediana). In ogni distribuzione di frequenza i valori delle osservazioni variano da un minimo ad un massimo attorno alla media. Il 95esimo percentile è il valore che sta sopra il 95% delle osservazioni e sotto il 5%. La deviazione standard (SD) descrive la variabilità di una singola osservazione. In una distribuzione simmetrica circa il 95% delle osservazioni è compresa entro due SD della media. Il coefficiente di variazione (CV) è una misura di dispersione relativa, espressa come percentuale dalla media. Ogni distribuzione di frequenza è descritta dalle stime dei "parametri" della popolazione da cui i dati sono stati estratti, e cioè dalle misure delle tendenze centrali della distribuzione (media, mediana e moda) e dalle misure di dispersione delle osservazioni attorno alla media (varianza, deviazione standard).

La distribuzione binomiale e la sua pdf sono descritte dai "parametri" n e p : $B(n, p)$. La media di $B(n, p)$ è $m = np$. La varianza di $B(n, p)$ è $S^2 = np(1 - p)$. La deviazione standard (SD) è la radice quadrata della varianza. Se campioniamo $n = 100$ individui da una popolazione in cui il genotipo G ha frequenza $p = 0.20$, ci aspettiamo di ottenere circa $m = np = 20$ G, con varianza $= np(1 - p) = 20 \times 0.8 = 16$, e con $SD = 4$.

Per esempio: le analisi di laboratorio hanno fornito una lista di genotipi, ottenuti identificando gli alleli che sono presenti in un campione di $n = 50$ individui al locus A (VNTR o STR). I risultati sono presentati in forma tabulare oppure come distribuzione di frequenza. Poniamo che nel campione analizzato siano presenti 11 individui che hanno il genotipo a_1a_2 . La frequenza di questo genotipo nel campione si calcola come proporzione: $11/50 = 0.22$. Questa frequenza è una stima corretta della frequenza del genotipo a_1a_2 nella popolazione? Se il genotipo $G = a_1a_2$ ha

frequenza $p = 0.22$, la probabilità di osservare $x = 11$ G su $n = 50$ individui campionati è data dalla pdf binomiale $B(50, p)$:

$$\Pr(x = 11|p) = [50! / 11!(50 - 11)!]p^{11}(1 - p)^{39} = 0.22$$

Se la probabilità di un evento è molto piccola, la distribuzione binomiale diventa corrispondente ad una **distribuzione di Poisson**:

$$\Pr(x|\lambda) = \lambda^{x-e} / x!$$

In questa distribuzione la media è uguale alla varianza: $m = S^2 = \lambda$. Se ogni evento ha più di due possibili risultati, la distribuzione diventa **multinomiale**.

Una distribuzione binomiale con $B(n, 0.5)$ è simmetrica. Se n è grande la distribuzione binomiale diventa **normale**: $N(m, S^2)$ (Fig. 43). Quando una distribuzione normale è continua, l'area compresa sotto la curva della distribuzione è = 1, cioè la probabilità che tutti gli eventi si verifichino è del 100%. Esistono infinite distribuzioni normali che corrispondono a tutte le possibili medie e varianze, ma esse possono essere standardizzate ed avere media = 0 e varianza = 1. La standardizzazione si ottiene trasformando i valori di x , appartenenti a $N(m, S^2)$, nella variabile $z = (x - m)/S$; $N(0, 1)$. Una distribuzione normale standardizzata ha il 95% dei suoi valori entro due SD dalla media. I valori di $z = +/- 1,96$ includono il 95% della distribuzione normale. Quindi, c'è una probabilità del 5% che una variabile standardizzata distribuita normalmente abbia valori > oppure < di 1,96, cioè al di fuori di due SD dalla media (Fig. 43). La pdf di una variabile binomiale standardizzata è:

$$z = (x - np) / \sqrt{np(1 - p)}$$

Questa formulazione consente di stimare le probabilità di eventi binomiali. Tuttavia, per poter applicare la distribuzione binomiale e le distribuzioni derivate alla genetica forense occorre stimare i parametri della distribuzione a partire dal campione. Non si tratta quindi di dedurre la probabilità di un evento data una distribuzione teorica di cui si conosce il valore dei parametri, ma, al contrario, di stimare il valore dei parametri della distribuzione di frequenza a cui appartengono gli eventi osservati.

In forma di pdf la probabilità di x è stata condizionata sul valore di p :

$$\Pr(x = n|p)$$

E' possibile stimare la probabilità di p , condizionando su un valore fisso di x :

$$L(p|x = n) = k(p)^x(1 - p)^{(n-x)}$$

In questa forma la funzione è detta: “likelihood” (verosimiglianza) di p dato $x = 11$, con $k =$ costante di proporzionalità che sostituisce il termine $[50! / 11!(50 - 11)!]$. La distribuzione di L è uguale alla distribuzione binomiale, ma la scala verticale dipende dal valore di k . In questo caso, la distribuzione di L ha un massimo a $p = 0.22$, vale a dire che 0.22 è il valore più verosimile del parametro p sulla base dei dati. Tramite il metodo della massima verosimiglianza (maximum likelihood; *ML*) si individua questo valore come stima del parametro. In generale, $L(p) = x/n$, se il parametro p da stimare appartiene ad una pdf binomiale. Quindi, $11/50 = 0.22$ è una buona stima della frequenza del genotipo G nella popolazione. Esistono situazioni più complicate in cui il parametro non appartiene ad una distribuzione binomiale e quindi deve essere stimato tramite *ML*.

Stima degli intervalli di confidenza

Gli intervalli di confidenza indicano la precisione delle stime dei parametri puntiformi, e sono calcolati in modo che, quando un esperimento è ripetuto più volte, il valore vero del parametro sia compreso all'interno dell'intervallo una pre-specificata percentuale di volte. Per esempio, un intervallo di confidenza del 95% assicura che il valore del parametro è compreso all'interno dell'intervallo nel 95% dei campioni. In genetica forense è importante calcolare gli intervalli di variazione delle frequenze alleliche: le frequenze degli alleli più rari possono essere determinate accuratamente solo esaminando centinaia, o migliaia di campioni, cosa che di solito non avviene. Quindi è necessario indicare gli intervalli di confidenza delle stime. Qualsiasi variabile standardizzata, distribuita normalmente con media $m = x/n$ e varianza $s^2 = np(1 - p)$, ha valori compresi fra $m \pm 1,96$ con il 95% delle probabilità. L'errore standard della media, che indica quanto varia la stima della media è: $sem = s/\sqrt{n}$. L'intervallo di confidenza che include il 95% dei valori della media di una variabile che, con probabilità p , si verifica x volte su un totale di n eventi è:

$$m \pm 1,96 \sqrt{p(1 - p)/n}$$

Questa formula è valida solo se il campione è sufficientemente grande, cioè se sia np che $n(1 - p) > 5$.

Le formule per gli intervalli di confidenza possono essere utilizzate per determinare le dimensioni del campione necessarie per ottenere un livello prefissato di precisione nelle stime:

$$n = (1.96 / d)^2 p(1 - p)$$

Occorre analizzare n individui, se si desidera ottenere una stima della frequenza di un allele che sia accurata con $\pm l$ con il 95% di confidenza e la frequenza attesa nella popolazione è p' .

Quando un locus presenta alleli multipli, gli intervalli di confidenza dei singoli alleli non equivalgono all'intervallo multiplo. Per ottenere una corretta stima simultanea degli intervalli di confidenza per alleli multipli, si applica la correzione di Bonferroni. Se si sono stimate le frequenze di k alleli ad un locus con una confidenza di $(1 - \alpha) \times 100\%$ (α = significatività = $1 - p$; di solito $p = 95\%$ o 99%), la correzione di Bonferroni è = $(1 - \alpha/k) \times 100\%$.

Test di significatività

Le procedure per calcolare i test di significatività si basano sul confronto fra i risultati attesi data un'ipotesi probabilistica (ipotesi nulla; H_0), ed i risultati osservati. Per dati categorici il test più semplice è il test del chi-quadrato:

$$\chi^2 = (O - E)^2 / E$$

con:

O = numero di osservazioni;

E = numero atteso di osservazioni data H_0

H_0 viene respinta se il valore di χ^2 è maggiore del valore di riferimento calcolato dalla distribuzione teorica di χ^2 , con M gradi di libertà (df) e al livello di significatività (α) prefissato. Ad esempio, nella distribuzione di χ^2 con 1 df , un valore > 3.84 occorre $< 5\%$ delle volte se l'ipotesi nulla è vera. Valori maggiori di 3.84 portano a respingere l'ipotesi nulla al livello di significatività del 5%, cioè con un valore di P del 5%. Il test di χ^2 può dare risultati errati quando il numero di eventi attesi è piccolo. In questi casi è meglio utilizzare i test esatti. I test esatti assumono che l'ipotesi nulla sia vera e calcolano la probabilità del risultato osservato o di valori più estremi (e quindi meno probabili) del risultato osservato. Se le probabilità di questi valori sono basse, l'ipotesi nulla è falsa. L'analisi statistica delle frequenze alleliche in genetica delle popolazioni ed in genetica forense consente di verificare se una popolazione di riferimento sia in HWI ed LE. Queste analisi possono essere effettuate utilizzando uno qualunque dei numerosi software di analisi statistica disponibili.

Stima delle frequenze alleliche e genotipiche

Lo scopo delle analisi di genetica forense è di verificare l'ipotesi che un determinato DNA fingerprinting sia associato univocamente ad un certo individuo, oppure che il DNA fingerprinting di un figlio sia derivato dai DNA fingerprinting dei due ipotetici genitori. Le analisi del DNA possono giungere alle seguenti conclusioni:

- non è possibile affermare con certezza se i campioni abbiano DNA fingerprinting identici (**risultati inconclusivi**). Questo può avvenire per una serie di diversi motivi: i campioni possono essere degradati, contaminati, la variabilità genetica può essere insufficiente. Le analisi possono essere ripetute con gli stessi o con differenti metodi nel tentativo di migliorare i risultati. I risultati di un'analisi inconclusiva non sono utilizzabili; è come se le analisi non fossero state fatte;
- i fingerprinting sono differenti e quindi devono essere originati da differenti individui (**esclusione**); una esclusione ha valore assoluto e non richiede ulteriori analisi o discussioni;
- i DNA fingerprinting sono corrispondenti (**inclusione**) e **possono** confermare le associazioni ipotizzate.

Nel terzo caso il problema diventa il seguente: qual è la significatività della corrispondenza? In altri termini: qual'è la probabilità che due differenti individui nella popolazione di riferimento abbiano per caso due DNA fingerprinting identici? Se, per ipotesi, i campioni sottoposti ad analisi genetica originassero da individui appartenenti ad una piccola popolazione, riproduttivamente isolata da molte generazioni, quindi con valori di inbreeding elevati e con variabilità genetica ridotta, due differenti individui potrebbero avere significative probabilità di condividere lo stesso DNA fingerprinting. In questo caso, una diagnosi di identità genetica (inclusione) potrebbe essere sbagliata.

Due campioni possono apparire geneticamente simile per i seguenti motivi:

- perché i due campioni provengono dallo stesso individuo;
- per coincidenza; i due campioni provengono da due individui che sono geneticamente simili per caso;
- per sbaglio: i due campioni provengono da due individui che sono geneticamente differenti, ma che sono risultati simili in conseguenza di errori commessi durante il percorso analitico (campioni identificati male, errori di analisi, metodi di laboratorio inadeguati, ecc.).

Come distinguere fra queste tre possibilità? Il controllo della qualità delle analisi condotte nel laboratorio di genetica forense può portare ad

escludere il terzo caso. Considerazioni di genetica delle popolazioni possono escludere il secondo caso. Se solo uno o pochi individui nella popolazione possono avere il DNA fingerprinting evidenziato nelle analisi, allora una coincidenza è molto improbabile. Se, al contrario, una parte della popolazione può condividere lo stesso DNA fingerprinting, allora due individui possono essere identici per coincidenza. Ad esempio, un falso genitore può avere per caso lo stesso genotipo del padre biologico. Il problema diventa: qual'è la probabilità che due individui della popolazione abbiano lo stesso DNA fingerprinting per caso? Oppure, ci si può chiedere: qual'è la probabilità di trovare quel particolare DNA fingerprinting se il padre presunto è il vero padre, in confronto con la probabilità di trovare lo stesso DNA fingerprinting se qualcun altro, ma non il padre presunto, è il vero padre? La risposta dipende dalla frequenza di quel particolare DNA fingerprinting nella popolazione di riferimento.

Per determinare la frequenza di un DNA fingerprinting è necessario analizzare un campione rappresentativo di individui della popolazione di riferimento, e contare il numero di volte in cui ogni genotipo, cioè ogni DNA fingerprinting, è presente. Se si esaminano loci con pochi alleli, genotipi identici appariranno molto frequentemente. Ad esempio, se un locus ha solo due alleli, nella popolazione ci saranno solo tre genotipi, ciascuno dei quali sarà probabilmente condiviso da molti individui. In questo caso sarà sufficiente esaminare un piccolo campione di individui per ottenere una stima precisa della frequenza dei tre genotipi. Se si esaminano loci con molti alleli, il numero dei possibili genotipi aumenta. In questo caso occorre esaminare un campione molto più ampio per ottenere stime precise delle frequenze dei genotipi. Quando il DNA fingerprinting è composto da loci multipli ipervariabili, il numero di possibili genotipi diventa elevatissimo ed è improbabile o addirittura impossibile trovarli tutti negli individui della popolazione. In questo caso la stima delle frequenze alleliche diventa difficile e la stima delle frequenze genotipiche diventa impossibile anche esaminando campioni molto grandi. Per esempio, se si esaminano minisatelliti umani con sistemi MLP, ogni locus può avere anche 50 alleli, che possono combinarsi e produrre 1.275 diversi genotipi. Se si esaminano quattro loci ciascuno con 50 alleli, il numero dei possibili genotipi sarà $(1.275)^4$, cioè circa 2,6 trillioni. Poiché esistono attualmente circa 6 miliardi di esseri umani, è evidente che la maggior parte di questi genotipi semplicemente non esiste. Perciò una stima diretta delle frequenze di questi genotipi non è possibile.

Occorre allora ottenere stime indirette delle frequenze (attese) dei genotipi. Queste stime si possono calcolare sulla base della teoria della

genetica delle popolazioni. Se assumiamo che una popolazione sia in HWE e in LE, allora è possibile stimare le frequenze genotipiche sulla base delle frequenze alleliche che sono state osservate nella popolazione. Poiché il numero totale di alleli presenti nella popolazione è comunque molto più piccolo del numero totale di genotipi, è molto più semplice stimare le frequenze alleliche. Le popolazioni reali hanno sempre dimensioni finite ed in alcuni casi possono essere anche molto piccole, gli accoppiamenti spesso non sono casuali, ma avvengono sulla base di certi criteri di scelta dei partners (ad esempio, negli umani, le scelte possono essere fatte sulla base del ceto, della religione, della etnia di provenienza; nella specie animali le scelte di coppia spesso sono fatte sulla base di caratteri morfologici che segnalano la fitness dei riproduttori). In altri casi può esserci migrazione e scambio di individui fra popolazioni differenziate. Molte popolazioni umane e molte “popolazioni” di animali riprodotti in cattività, sono in realtà misture di individui che originano da popolazioni diverse e geneticamente differenziate. In questi casi, le popolazioni non sono insieme indistinti di individui, ma sono stratificate, cioè sono strutturate in sottogruppi che possono essere fra di loro geneticamente differenziati. E’ probabile che una popolazione strutturata e suddivisa in sottopopolazioni non sia in HWE. Ovviamente, gli individui che appartengono a ciascun sottogruppo saranno fra di loro meno geneticamente differenziati di quanto ci si potrebbe attendere se la popolazione non fosse strutturata. Gli individui di un sottogruppo possono non essere in LE.

Non necessariamente la stima delle frequenze genotipiche in popolazioni strutturate presenta dei problemi pratici. Infatti:

- le deviazioni da HWE ed LE possono essere piccole e non invalidare le stime delle frequenze genotipiche calcolate utilizzando stime delle frequenze alleliche;
- gli effetti della struttura delle popolazioni sono prevedibili e possono essere corretti matematicamente. In una popolazione strutturata ci si aspetta di trovare più omozigoti e meno eterozigoti, di quanti ne sono attesi in una popolazione mendeliana. L’effetto della struttura di popolazione su LE è di aumentare la correlazione fra alcuni loci e di diminuirla fra altri. Dopo aver raccolto dati empirici sulla variabilità genetica di una popolazione, si possono quantificare le deviazioni da HWE e LE, si possono quindi correggere gli errori di stima delle frequenze genotipiche. E’ possibile ricostruire i sottogruppi presenti in una popolazione, anche senza conoscerne la composizione a priori.

Stima delle frequenze alleliche a loci codominanti (VNTR o STR)

Mendel derivò le sue leggi da osservazioni sulle frequenze di caratteri fenotipici la cui espressione è controllata da relazioni di dominanza/recessività. Gli alleli dominanti impediscono l'espressione degli alleli recessivi, che non possono essere direttamente osservati. Esistono metodi statistici che consentono di stimare le frequenze alleliche anche a loci con alleli dominanti/recessivi. Tuttavia, l'analisi diretta del DNA consente di individuare entrambi gli alleli presenti ai loci eterozigoti (ad eccezione dei DNA fingerprinting multilocus, in cui l'identificazione dei singoli alleli è problematica), che possono essere trattati come loci codominanti. L'analisi dei loci codominanti consente di calcolare le frequenze alleliche direttamente dal conteggio dei genotipi presenti nel campione, tramite la formula:

$$p(a_1) = p(a_1a_1) + 1/2 p(a_xa_1)$$

con:

$p(a_1)$ = frequenza dell'allele a_1

$p(a_1a_1)$ = frequenza del genotipo omozigote a_1a_1

$p(a_xa_1)$ = frequenza dei genotipi eterozigoti che contengono l'allele a_1

Se un locus presenta solo 2 alleli, a_1 e a_2 , ci sarà un solo genotipo eterozigote: a_1a_2 . Ma se il locus presenta più di due alleli (alleli multipli) ci saranno due o più possibili genotipi eterozigoti. Gli omozigoti contengono due copie dello stesso allele, mentre gli eterozigoti contengono solo una copia di ogni allele. Perciò è necessario dividere per 2 le frequenze degli eterozigoti.

In un sistema con due alleli codominanti, questa formulazione è equivalente alla formulazione seguente:

$$p(a_1) = [2(a_1a_1) + (a_2a_1)]/2n = p$$

$$p(a_2) = 1 - p(a_1) = q$$

con: (a_1a_1) = numero degli omozigoti nel campione; (a_2a_1) = numero degli eterozigoti nel campione; n = numero totale degli individui analizzati.

L'errore standard delle stime è: $\sqrt{(pq/2n)}$

Stima delle frequenze alleliche in minisatelliti analizzati con sistemi MLP

Nei DNA fingerprinting multilocus, l'identificazione dei singoli alleli è problematica. In pratica, due frammenti sono identificati come "corri-

spondenti” se hanno un peso molecolare che è determinato entro 3 unità di deviazione standard in ogni senso di migrazione. In sistemi elettroforetici ben calibrati, ogni frammento ha una deviazione standard corrispondente a circa lo 0.6% del suo peso molecolare. Quindi due frammenti sono corrispondenti se hanno lo stesso MW \pm 2.0 - 2.5% di scostamento dal loro MW medio. Due frammenti corrispondenti vengono assegnati allo stesso “allele”. In genetica forense umana gli “alleli” di MW simile vengono raggruppati entro lo stesso bin. Ogni bin deve contenere almeno 5 alleli diversi. Poi si calcolano le frequenze dei bin a cui appartengono gli “alleli”. In genetica forense di specie CITES quasi sempre mancano adeguati data-base delle popolazioni di riferimento. Perciò gli alleli vengono identificati sulla base dell’intervallo di variazione dei singoli frammenti. Ogni allele così definito corrisponde ad un bin. Le frequenze alleliche e le frequenze dei bin nelle popolazioni di riferimento sono equivalenti. La frequenza di ogni frammento è calcolata semplicemente come proporzione: numero dei campioni in cui il frammento è presente / il totale dei campioni analizzati.

Stima delle frequenze genotipiche a sistemi multilocus

La legge di Hardy-Weinberg consente di calcolare le frequenze genotipiche attese ad un locus singolo, utilizzando le frequenze alleliche stimate nel campione:

- genotipi omozigoti: $p(a_1a_1) = p(a_1)^2$: la frequenza del genotipo omozigote a_1a_1 è uguale al quadrato della frequenza dell’allele a_1
- genotipi eterozigoti: $p(a_xa_1) = 2p(a_1)p(a_x)$: la frequenza del genotipo eterozigote che contiene l’allele a_1 è uguale al doppio del prodotto delle frequenze degli alleli

Le frequenze genotipiche stimate tramite le frequenze alleliche sono corrispondenti alle frequenze genotipiche della popolazione solo se la popolazione è in HWE. In una popolazione in equilibrio, le frequenze alleliche e genotipiche non cambiano da una generazione all’altra. Le forze evolutive (selezione, mutazione, migrazione) cambiano le frequenze alleliche da una generazione all’altra. In una popolazione che evolve, le frequenze genotipiche attese non sono stimabili direttamente tramite le frequenze alleliche e la legge di Hardy-Weinberg.

I profili dei DNA fingerprinting usati in genetica forense sono quasi sempre determinati usando sistemi multilocus. E’ possibile stimare la frequenza di un genotipo multilocus come prodotto delle frequenze degli

alleli che sono presenti a ciascuno dei loci che compongono il profilo. Se il sistema multilocus è composto dai loci A (con gli alleli a_1 e a_2), B (con gli alleli b_1 e b_2), e C (con gli alleli c_1 e c_2), la frequenza del genotipo multilocus è calcolata tramite la “regola del prodotto” che generalizza la legge di Hardy-Weinberg:

$$p(ABC) = 2^H p(a_1)p(a_2)p(b_1)p(b_2)p(c_1)p(c_2)$$

Cioè: la frequenza del genotipo ABC è uguale al prodotto delle frequenze degli alleli che lo compongono, moltiplicato per 2^H , con H = numero di loci che sono eterozigoti nel profilo. Questa stima è corretta se e solo se la segregazione fra tutti gli alleli presenti nel profilo è indipendente, cioè se la popolazione è in HWE e se tutti gli alleli sono in LE. Queste assunzioni sono difficilmente soddisfatte (ed è anche problematico verificarle) in popolazioni reali di dimensioni finite. Tuttavia, spesso le deviazioni da HWE e LE non sono tali da invalidare la stima delle frequenze genotipiche.

Effetto della stratificazione. Se una popolazione (non in HWE) è composta da sottopopolazioni differenziate, ciascuna delle quali è in HWE, le frequenze genotipiche possono essere calcolate dalle frequenze alleliche nella popolazione totale introducendo il fattore di correzione q . Se sono note le frequenze alleliche nella sottopopolazione, allora le frequenze genotipiche si calcolano esattamente come sopra senza necessità di introdurre alcuna correzione.

Intervalli di confidenza per genotipi multilocus. Se la dimensione del campione è grande e se la frequenza del genotipo al locus P non è molto più piccola di 0,5, allora la distribuzione binomiale $B(n, P)$ di P tende ad una distribuzione normale e l'intervallo di confidenza al 95% per P si calcola come:

$$P \pm 1,96\sqrt{P(1 - P) / n}$$

Tuttavia gli intervalli di confidenza dei genotipi ai singoli loci sono di utilità limitata, poiché siamo interessati a definire le caratteristiche delle distribuzioni di frequenza dei genotipi multilocus. Gli intervalli di confidenza multilocus richiedono l'applicazione di statistiche complesse, e sono calcolabili tramite appositi software. Se la frequenza del genotipo multilocus P è piccola, spesso gli intervalli di confidenza possono essere $(P/10, 10P)$. Quindi è impossibile stimare con precisione piccole frequenze in piccoli campioni.

PROBABILITÀ

I calcoli che consentono di ottenere la stima delle frequenze genotipiche si basano sulle regole del calcolo delle probabilità. Qual è la probabilità che due campioni che hanno profilo genetico simile appartengano allo stesso individuo? Qual è la probabilità che un padre presunto sia il padre biologico quando tutti gli alleli del figlio sono stati identificati nella madre e nel padre presunto? La teoria delle probabilità è utilizzata per rispondere a queste domande. Il calcolo delle probabilità ha lo scopo di quantificare l'incertezza, cioè di assegnare un valore di probabilità ad un evento incerto tramite procedure di stima (analisi statistica). L'incertezza deriva dalla complessità, che rende impossibile controllare tutte le connessioni di causa-effetto e disporre di tutte le informazioni necessarie per comprendere i processi reali. La probabilità è una misura dell'incertezza espressa con un numero che varia fra 0 ed 1. Ci sono diversi concetti di probabilità. Le probabilità sono determinate soggettivamente, oggettivamente, oppure empiricamente. Le probabilità soggettive sono basate sull'esperienza che consente di assegnare stime di verosimiglianza agli eventi. Le probabilità oggettive sono basate su dati sperimentali che consentono di calcolare la frequenza di un evento. Le probabilità empiriche sono basate sulle informazioni acquisite dall'analisi di dati già disponibili.

Teoria frequentistica della probabilità (su cui si basa la statistica classica)

La probabilità p di un evento H dipende dal numero di volte (n) in cui l'evento si realizza sul totale delle prove (N). La probabilità p di H corrisponde quindi alla sua frequenza:

$$pH = n(H) / N$$

Questa definizione si basa sull'assunzione che tutti gli eventi siano egualmente probabili (equiprobabili). Ovviamente, se il numero delle prove è piccolo, la stima di pH sarà incerta. La "legge dei grandi numeri" assicura che, ripetendo le prove un numero molto alto di volte (tendente all'infinito), pH sarà determinata con precisione. Nella statistica classica il valore di pH può essere determinato solo sperimentalmente, perché non ci sono leggi di natura universale che garantiscano l'equiprobabilità degli eventi. Anche nel lancio della moneta non ci sono leggi della fisica che garantiscano che p "testa" = p "croce" = 50%. Se lanciamo una moneta poche volte, certamente $p_t \neq p_c$ (t = testa; c = croce). Se lanciamo una

moneta un numero molto alto (tendenzialmente infinito) di volte, $p_t = p_c$ (ma questo deve essere verificato sperimentalmente). Il risultato di un evento (variabile casuale) può avere due valori (vero o falso) o più valori numerici. La statistica classica è poco utilizzabile per determinare l'incertezza di eventi che non sono sperimentabili. Ad esempio: qual è la probabilità che domani piova? Qual è la probabilità che il genotipo di questo campione corrisponda al genotipo di questo individuo?

Teoria soggettivistica della probabilità (statistica Bayesiana)

La probabilità p è una stima della verosimiglianza che l'evento H si verifichi. Possiamo avere convinzioni (soggettive) o informazioni (oggettive, anche se non esattamente quantificabili) che un evento possa verificarsi più o meno frequentemente. Le circostanze, che non corrispondono alle frequenze determinate sperimentalmente, consentono di assegnare una probabilità all'evento. Da questo punto di vista, le probabilità assegnate ad un evento sono condizionali, cioè valgono solo in presenza di determinate circostanze. Per esempio: $p_t = p_c = 0.5$, se e solo se sappiamo che una moneta ha un lato con una testa ed un lato con una croce, che i due lati hanno lo stesso peso, che la moneta è lanciata in modo da randomizzare gli eventi (randomizzazione: procedure di massimizzazione dell'incertezza, della casualità; massima incertezza: quando tutti gli eventi possibili hanno esattamente la stessa probabilità di realizzarsi, cioè hanno la stessa verosimiglianza), ecc. Se sappiamo che $p_t = p_c = 0.5$, allora ci aspettiamo che in realtà $p_t = p_c$ se lanciamo la moneta un numero di volte molto alto. La previsione del risultato dei lanci di una moneta non dipende da una legge fisica, ma dalle condizioni associate al lancio.

In che modo determinare la probabilità di un evento incerto di cui non possiamo sperimentare la frequenza? Cioè: qual è la $\Pr(H|E)$ = la probabilità che si verifichi l'evento H data l'evidenza E ? I fattori che condizionano $\Pr E$ possono essere molteplici, per esempio: $\Pr(H|S, C, I)$, dove S , C e I indicano i dati (cioè le osservazione quantificabili) e le informazioni (cioè i dati non esattamente quantificabili) che sono rilevanti per determinare $\Pr H$. I dati e le informazioni costituiscono l'evidenza. Se riteniamo che tutte le probabilità siano condizionali, allora $\Pr(H|E) = \Pr(E)$, e le due notazioni sono equivalenti.

Le regole del calcolo delle probabilità

- Prima legge della probabilità: i valori di probabilità vanno da 0 a 1:
 $0 > \Pr(H|E) < 1$.

La probabilità complementare di $\Pr(H|E)$ è: $1 - \Pr(H|E)$. Se E avviene (è avvenuto, avverrà certamente), allora $\Pr = 1$, e la sua probabilità complementare sarà $1 - 1 = 0$.

- Seconda legge della probabilità: due eventi sono reciprocamente esclusivi se il verificarsi dell'uno ($t = \text{testa}$) esclude il verificarsi dell'altro ($c = \text{croce}$; il risultato del lancio di una moneta è testa \bullet è croce). La probabilità che si verifichi l'uno o l'altro dei due eventi mutualmente esclusivi è data dalla somma delle loro rispettive probabilità (regola dell'addizione): $\Pr(t \bullet c|E) = \Pr(t|E) + \Pr(c|E) = 1$.

La probabilità complementare di un evento esclusivo è:

$$\Pr(t|E) = 1 - \Pr(c|E).$$

I due eventi sono esaustivi.

- Terza legge della probabilità: due eventi sono indipendenti se il verificarsi dell'uno non influenza il verificarsi dell'altro. La probabilità che si verifichino entrambi (l'uno e l'altro) è data dal prodotto delle loro rispettive probabilità (regola del prodotto): $\Pr(A \text{ e } B|E) = \Pr(A|E)\Pr(B|A, E)$.

Naturalmente, le formulazioni seguenti sono equivalenti:

$$\Pr(A \text{ e } B) = \Pr(A)\Pr(B|E) = \Pr(B)\Pr(A|E) = \Pr(A)\Pr(B)$$

In questo caso i due eventi sono statisticamente indipendenti (condizionali su E). Due eventi possono essere indipendenti assumendo un'ipotesi, ma dipendenti assumendone un'altra.

Esempio: il locus A ha l'allele a con frequenza $p = 0.2$; il locus B ha l'allele b con frequenza $p = 0.3$. I due loci sono indipendenti. La probabilità di ottenere un genotipo con entrambi gli alleli a e b si calcola tramite la regola del prodotto: $\Pr(a \text{ e } b) = \Pr(a)\Pr(b) = 0.2 \times 0.3 = 0.06$.

Esempio: una popolazione è composta da due sottopopolazioni: 25% A e 75% C . Il genotipo G è presente nel 4,8% degli individui della popolazione A . Qual'è la probabilità che una persona scelta a caso dalla popolazione totale provenga dalla sottopopolazione A ed abbia il genotipo G ? La probabilità è: $\Pr(A \text{ e } G) = \Pr(A)\Pr(G|A) = 0.25 \times 0.048 = 0.012$.

- Eventi parzialmente associati. Due eventi possono essere solo parzialmente indipendenti ed avere qualche cosa in comune. In questo caso:

$\Pr(A \text{ o } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ e } B)$, cioè la probabilità di un'evento dipende in parte anche dal verificarsi dell'altro evento. La probabilità che entrambi gli eventi si realizzino è:

$\Pr(A \text{ e } B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$, dove $\Pr(A|B)$ è la probabilità condizionale che A si realizzi se B si è realizzato.

Esempio: la probabilità di ottenere al locus A un genotipo con solo l'allele a o solo l'allele b o con entrambi gli alleli a e b è: $\Pr(a \text{ o } b) = \Pr(a) + \Pr(b) - \Pr(a \text{ e } b) = 0.2 + 0.3 - 0.06 = 0.44$.

- La legge della probabilità totale. Se due eventi A e B sono mutualmente esclusivi ed esaustivi ($B = 1 - A$), la probabilità di un evento H che dipende da A e B è:

$$\Pr(H) = \Pr(H|A)\Pr(A) + \Pr(H|B)\Pr(B)$$

Esempio: una popolazione è composta da tre sottopopolazioni: 83,47% A, 12,19% B, 4,34% C. Il genotipo G è presente nelle tre sottopopolazioni con le frequenze 0,013, 0,045 e 0,039, rispettivamente. Qual è la probabilità di trovare G in un individuo preso a caso dalla popolazione totale? La Pr è $= 0,013 \times 83,47 + 0,045 \times 12,19 + 0,039 \times 4,34 = 0.018$.

Il teorema di Bayes

Abbiamo a disposizione 52 carte da gioco e vogliamo calcolare la probabilità di estrarre un Re. Ci sono 4 Re in ogni mazzo di carte, così la probabilità di estrarre un Re è $4/52$. Ora vogliamo calcolare la probabilità di estrarre una carta Rossa che sia anche un Re. Cioè: se la carta è un Re, qual è la probabilità (condizionale) che sia Rossa: $\Pr(\text{carta Rossa}|\text{Re})$. Un mazzo è composto dal 50% di carte Rosse e dal 50% di carte Nere, quindi $\Pr(\text{carta Rossa}|\text{Re}) = 2/4 = 0.5$. La probabilità inversa è la probabilità di estrarre un Re in un mazzo di carte che siano tutte Rosse: $\Pr(\text{Re}|\text{carta Rossa}) = 2/26$.

Il teorema di Bayes (formulato dal Reverendo Thomas Bayes; 1702 – 1761) mette in relazione queste due probabilità. Identifichiamo con: $A =$ carta Rossa, $B =$ Re, il teorema di Bayes stabilisce che:

- $\Pr(B|A) = \Pr(A|B)[\Pr(B) / \Pr(A)]$

In questa formulazione il teorema di Bayes richiede la conoscenza dei valori di due probabilità non-condizionali note a-priori, che sono: $\Pr(A) = \Pr(\text{carta Rossa}) = 50\% = 26 \text{ carte Rosse} / 52 \text{ carte in un mazzo}$; $\Pr(B) = 4 \text{ Re} / 52 \text{ carte in un mazzo}$: Occorre inoltre conoscere il valore di una probabilità condizionale nota a-priori, che è: $\Pr(A|B) = \Pr(\text{carta Rossa}|\text{condizionata al fatto che sia un Re}) = 2/4$. Applicando il teorema di Bayes possiamo calcolare la probabilità di estrarre un Re condizionata al fatto che la carta sia Rossa:

$$\Pr(B|A) = (2/4)[(4/52) / (26/52)] = 2/26$$

La probabilità condizionale ignota (a-posteriori) si può calcolare dalla probabilità condizionale e dalle due probabilità non-condizionali note a-priori. In questo caso la probabilità a-posteriori era calcolabile direttamente, sapendo che il 50% delle carte di un mazzo sono Rosse, inclusi due Re su quattro. Ma esistono innumerevoli casi in cui le probabilità a-posteriori non sono note, mentre sono note, o sono ipotizzabili, le probabilità a-priori. Il teorema di Bayes consente di stimare i valori di probabilità e di aggiornarli sulla base delle informazioni fornite dall'evidenza. Disponiamo di una stima a-priori della probabilità di realizzazione dell'evento, che viene combinata con la probabilità condizionale (likelihood) dell'evidenza nel caso in cui l'evento si sia realizzato, per ottenere la probabilità aggiornata (a-posteriori) dell'evento data l'evidenza. L'informazione nota prima dell'evento è detta "probabilità a-priori"; l'informazione nota dopo l'evento è detta "probabilità a-posteriori".

Il teorema di Bayes può essere espresso nella forma seguente:

$$\bullet \Pr(Hp|E, I) / \Pr(Hd|E, I) = [\Pr(E|Hp, I) / \Pr(E|Hd, I)] \times [\Pr(Hp|I) / \Pr(Hd|I)]$$

con: E = evidenze; I = altre informazioni che condizionano la Pr dell'evento Hp e la Pr dell'evento alternativo Hd .

In questo modo il teorema di Bayes è scritto in forma di pronostico. Se sappiamo che la probabilità di H è $\Pr(H)$, il pronostico O in favore di H è:

$$O(H) = \Pr(H) / \{1 - \Pr(H)\}$$

Sappiamo che $\Pr(H) + \{1 - \Pr(H)\} = 1$; $O(H)$ può variare da 0 (se H è falso) ad infinito (se H è vero). Quando il pronostico è negativo (esempio: 1/5) di solito si usa il pronostico a sfavore, corrispondente all'inverso del pronostico a favore (esempio: 5 a 1). Quando i pronostici a favore e a sfavore sono uguali, cioè = 0.5, allora si dicono pari (evens). La conversione dei pronostici in probabilità è:

$$\Pr(H) = O(H) / 1 + O(H)$$

Esempio: se $O(H) = 1/5$, $\Pr(H) = 1/5 / 1 + 1/5 = 0.17$

Il teorema di Bayes espresso nella forma di pronostico consente di calcolare le probabilità a-posteriori di un evento [pronostico a-posteriori = $\Pr(Hp|E, I) / \Pr(Hd|E, I)$], come prodotto di due rapporti:

- il rapporto fra le probabilità condizionali dell'evidenza dati l'evento e l'evento alternativo, che chiamiamo il rapporto di verosimiglianza (likelihood ratio) = $[\Pr(E|Hp, I) / \Pr(E|Hd, I)]$
- il rapporto fra le due probabilità note a-priori = $[\Pr(Hp|I) / \Pr(Hd|I)]$.

APPLICAZIONI DELLA STATISTICA BAYESIANA ALLA GENETICA FORENSE

L'approccio Bayesiano è esplicitamente o implicitamente sottinteso alla soluzione dei problemi di genetica forense (vedi, per esempio il rapporto tecnico U.S.A. National Research Council - NRC, 1966). Nel presentare le applicazioni del modello Bayesiano abbiamo seguito il testo di Evett and Weir (1998).

Identificazione

Supponiamo che la guardia forestale abbia ritrovato i resti di una carcassa di cervo, apparentemente ucciso in periodo di caccia chiusa (oppure in un'area protetta), ed abbia poi confiscato porzioni di carne conservata nel congelatore di un ipotetico bracconiere. Sulla base di una serie di informazioni la guardia forestale ritiene che l'ipotetico bracconiere sia verosimilmente il responsabile dell'uccisione del cervo. Vengono prelevati campioni biologici dalla carcassa del cervo (il corpo del reato) e dalla carne congelata (il sospetto). Vengono determinati i DNA fingerprinting di questi campioni. I campioni presentano profili di DNA fingerprinting identici. Il problema è di stabilire se i due campioni appartengano allo stesso individuo, e se, come conseguenza, l'ipotetico bracconiere possa venire accusato del reato. Ovviamente, qualunque sia la risposta fornita dalle procedure di analisi genetica, essa non costituirà prova che l'accusato abbia commesso effettivamente il reato. Il sospetto potrebbe fornire spiegazioni plausibili per avere avuto in qualche modo ed in buona fede la carne sequestrata nel congelatore.

In una prospettiva Bayesiana si pone il problema di valutare due ipotesi H alternative, mutualmente esclusive ed esaustive:

- H_p : la carne congelata **appartiene** alla carcassa di cervo
- H_d : la carne congelata **non appartiene** alla carcassa di cervo

L'evidenza genetica G è costituita dal DNA fingerprinting:

- G_s : il DNA fingerprinting della carne congelata (**sospetto**);

- G_c : il DNA fingerprinting della carcassa (**corpo del reato**);

In questo caso le analisi di laboratorio dicono che: $G_s = G_c$.

Si devono poi valutare le informazioni di tipo non genetico I , cioè tutte le altre informazioni che possono sostenere l'accusa (H_p).

Prima dell'analisi del DNA la probabilità di H_p era condizionata solo da I : $\Pr(H_p|I)$. Dopo l'analisi del DNA la probabilità di H_p è condizionata da G_s , G_c e I : $\Pr(H_p|G_s, G_c, I)$. Per stimare $\Pr(H_p)$ è necessaria almeno un'ipotesi alternativa H_d . Per stimare la probabilità delle due ipotesi alternative possiamo usare il teorema di Bayes, espresso in forma di pronostici. Dobbiamo quindi valutare:

- i pronostici a-priori in favore di H_p : $\Pr(H_p|I) / \Pr(H_d|I)$;
- i pronostici a-posteriori in favore di H_p : $\Pr(H_p|G_s, G_c, I) / \Pr(H_d|G_s, G_c, I)$.

I pronostici a-priori devono essere noti prima. I pronostici a-posteriori possono essere calcolati usando il teorema di Bayes. Definiamo l'evidenza $E = (G_s, G_c)$:

$$\Pr(H_p|E, I) / \Pr(H_d|E, I) = [\Pr(E|H_p, I) / \Pr(E|H_d, I)] \times [\Pr(H_p|I) / \Pr(H_d|I)]$$

Questa formulazione significa chiedersi:

- qual è la probabilità $\Pr(E|H_p, I)$ dell'evidenza (DNA fingerprinting) ammesso che H_p sia vera?
- qual è la probabilità $\Pr(E|H_d, I)$ dell'evidenza (DNA fingerprinting) ammesso che H_d sia vera?

Il rapporto fra le due probabilità a-posteriori è il likelihood ratio $LR = \Pr(E|H_p, I) / \Pr(E|H_d, I)$. I pronostici a-posteriori si calcolano moltiplicando i pronostici a-priori per LR .

$$LR = \Pr(E|H_p, I) / \Pr(E|H_d, I) = \Pr(G_s, G_c|H_p, I) / \Pr(G_s, G_c|H_d, I)$$

Applicando la terza legge della probabilità si può espandere LR , che diventa:

$$LR = \Pr(G_d|G_s, H_p, I) / \Pr(G_d|G_s, H_d, I) \times \Pr(G_s|H_p, I) / \Pr(G_s|H_d, I)$$

I termini: $\Pr(G_s|H_p, I)$ e $\Pr(G_s|H_d, I)$ indicano le probabilità di osservare il genotipo G_s indipendentemente dal genotipo G_c . Quindi: $\Pr(G_s|H_p, I) = \Pr(G_s|H_d, I)$, poiché la verosimiglianza delle due ipotesi alternative non fornisce alcuna informazione sulla verosimiglianza del genotipo G_s . Quindi:

$$LR = \Pr(G_d|G_s, H_p, I) / \Pr(G_d|G_s, H_d, I) \times 1 = \Pr(G_d|G_s, H_p, I) / \Pr(G_d|G_s, H_d, I)$$

Poiché è certo che il genotipo è G_c nel caso in cui H_p è vera, allora: $\Pr(G_d G_s, H_p, I) = 1$, quindi:

$$LR = 1 / \Pr(G_d G_s, H_d, I)$$

Quello che resta da fare è assegnare le probabilità di G_c nel caso in cui il campione appartenesse ad un altro individuo e non al cervo trovato ucciso. Questa probabilità dipende da I (le circostanze). Se assumiamo che G_s non influenza l'incertezza su G_c , dato che essi nell'ipotesi appartengono a due individui diversi (assunzione che in certi casi può essere falsa, ad esempio se i due campioni provengono da individui imparentati), allora:

$\Pr(G_d G_s, H_d, I) = \Pr(G_d H_d, I)$, e quindi:

$$LR = 1 / \Pr(G_d H_d, I)$$

In che modo assegnare un valore al denominatore? Qual è la probabilità di osservare il genotipo $G_c = G_s = G$, se i due campioni analizzati non appartengono allo stesso individuo? La risposta dipende interamente da I . In questo caso occorre identificare il gruppo di individui da cui G proviene, cioè la popolazione da cui proviene il cervo. Occorre ottenere informazioni genetiche su un campione rappresentativo di individui della popolazione di riferimento, ed usare queste informazioni per ricavare stime sui parametri della popolazione, usando i metodi di calcolo della statistica. Assumiamo di aver fatto tutto questo e di sapere che il genotipo G è presente nella popolazione con una frequenza P , che corrisponde alla probabilità del denominatore. Allora:

- $LR = 1 / P$

Per esempio, se $P = 0.01 = 1 / 100$, allora $LR = 100$, il che significa che: "l'evidenza è 100 volte più probabile se il campione di carne ed il campione prelevato dalla carcassa appartengono allo stesso cervo di quanto non lo sia se essi appartenessero a due distinti cervi (non imparentati) provenienti dalla medesima popolazione". Questa conclusione si basa sulle informazioni che la condizionano: se i due campioni provengono da due individui diversi, essi sono o non sono imparentati? Come definire la popolazione da cui proviene il campione sospetto? E' una popolazione geneticamente omogenea oppure è composta da sottopopolazioni? Ogni individuo appartenente alla popolazione ha la stessa probabilità di essere ucciso? Le assunzioni utilizzate per condizionare la stima di probabilità devono essere rese esplicite, e, se nuove circostanze lo richiedono, devono essere modificate. Operativamente si arriva alla stessa conclusione ($LR = 1$

/ P) utilizzando un modello frequentista, ma in questo caso le assunzioni sottostanti non sono rese esplicite, e quindi non è possibile esplicitare gli elementi soggettivi che condizionano le stime di probabilità.

I tre principi dell'interpretazione delle evidenze (Evet and Weir, 1998)

- Per valutare l'incertezza di ogni ipotesi è necessario considerare almeno un'ipotesi alternativa.
- L'interpretazione delle evidenze scientifiche si basa sulla risposta alla domanda: "Qual è la probabilità di osservare l'evidenza data l'ipotesi"?
- L'interpretazione delle evidenze è condizionata non solo dalle ipotesi alternative, ma anche dalle circostanze nell'ambito delle quali l'evidenza deve essere valutata.

Eventuali informazioni (I) sulla dimensione e sull'isolamento della popolazione possono condizionare significativamente l'interpretazione dell'evidenza. Supponiamo, per esempio, che il cervo ucciso ed il campione sospetto derivino da una piccola popolazione isolata, composta da N individui, che non riceve immigranti da tempo. I due DNA fingerprinting sono identici e presentano il genotipo G . Il genotipo G ha frequenza P nella popolazione, cioè la probabilità che un cervo preso a caso dalla popolazione abbia genotipo G è $= P$. La probabilità a-priori che ogni cervo della popolazione sia stato ucciso è $\Pr(Hp|I) = 1 / N$, e $LR = 1 / P$.

La probabilità a-posteriori è:

$$\Pr(Hp|E, I) / \Pr(Hd|E, I) = (1 / P) \times (1 / N) = 1 / NP$$

che è verosimile se ogni individuo della popolazione ha la stessa probabilità di identità P , ma non lo è se teniamo conto delle possibili relazioni di parentela in una piccola popolazione. Allora è conveniente modificare la precedente formulazione per la stima delle probabilità a-posteriori come segue:

$$\Pr(Hp|E, I) / \Pr(Hd|E, I) = 1 / \sum_i w_i P_i$$

dove: $i = 1, N$; w_i sono le probabilità - variabili - che ogni individuo i della popolazione di dimensione N , abbia lo stesso genotipo del cervo ucciso, indipendentemente dalle evidenze fornite dal DNA, relativamente alle probabilità del genotipo del corpo del reato; P_i sono le probabilità individuali di identità col genotipo G . Se tutti i w_i sono $= 1$ e tutti i P_i sono $= 1$, allora le probabilità a-posteriori equivalgono a $1 / NP$, come sopra. Ma se sappiamo che esistono due cervi che sono fratelli, allora la probabilità che essi abbiano lo stesso genotipo $G = 1 / 4$. Assumendo che tutte le altre probabilità di identità individuale siano $= 1 / 100$, la probabilità a posteriori diventa:

$$1 / (1/4 + 1/100) = 4.$$

Dalla conclusione: “l’evidenza è 100 volte più probabile se il campione di carne ed il campione prelevato dalla carcassa appartengono allo stesso cervo di quanto non lo sia se essi appartenessero a due distinti cervi (non imparentati) provenienti dalla medesima popolazione”, possono derivare le seguenti estrapolazioni che non sono giustificate:

- la probabilità che i due campioni provengano da due distinti individui è 1 su 100, quindi c’è il 99% di probabilità che derivino dallo stesso individuo
- se la popolazione è composta da 1.000 individui e $P = 0.01$, allora ci aspetta di trovare 10 individui con lo stesso DNA fingerprinting; questo non vuol dire che ognuno di questi 10 individui ha la stessa probabilità (1 / 10) di essere stato ucciso
- se ci si aspetta di trovare un individuo con un DNA fingerprinting che ha $P = 0.01$ in una popolazione di 100 individui, non è detto che esso sia il cervo ucciso.

Probabilità di esclusione

Ogni sistema di marcatori usato in genetica forense deve avere sufficiente variabilità per consentire l’individualizzazione dei campioni, ma anche sufficiente stabilità per non introdurre mutazioni nell’ambito di una generazione (dai genitori ai figli). La probabilità di individualizzazione, cioè di identificare un genotipo che sia unico nella popolazione di riferimento, aumenta con l’aumentare del numero dei loci che sono usati nel sistema multilocus. Allo stesso modo aumenta la probabilità di esclusione. In un sistema codominante con k alleli, ciascuno dei quali con frequenza p_i nella popolazione ($i = 1, k$), la probabilità di esclusione media PE è data dalla formula:

$$PE = \sum^k [p_i(1 - p_i)^2 + \sum_i \sum_{j(i>j)} p_i p_j \{ (1 - p_i)^3 + (1 - p_j)^3 + (p_i + p_j)[1 - (p_i + p_j)]^2 \}]$$

Nell’analisi dei test di paternità le probabilità di esclusione ad ogni locus con due alleli codominanti dipende dalle frequenze alleliche e dagli alleli paterni che sono presenti nei genotipi di ogni figlio. I test di esclusione di paternità si usano per escludere che un padre presunto sia il padre biologico. L’esclusione si può fare direttamente quando il padre presunto non presenta gli alleli paterni del figlio ad almeno uno dei loci analizzati, oppure, se gli alleli paterni non possono essere definiti, egli non ha nessuno degli alleli del figlio. L’allele paterno ad ogni locus è identificato come quell’allele che il figlio non riceve dalla madre, ed è

facilmente identificabile ad eccezione dei casi in cui madre e figlio sono eterozigoti per gli stessi due alleli. In questo caso, se il locus ha solo due alleli, nessun maschio può essere escluso a questo locus. E' possibile calcolare le probabilità di esclusione in relazione ai genotipi dei tre individui coinvolti nel test: la madre, il figlio ed il padre presunto.

La tabella 3 riporta le probabilità di esclusione dei genotipi paterni per ogni coppia di genotipi madre-figlio per sistemi codominanti con due alleli. In questi casi, la probabilità di esclusione media è:

$$PE = pq(1 - pq)$$

La tabella 4 riporta le probabilità di esclusione dei genotipi paterni per ogni coppia di genotipi madre-figlio per sistemi codominanti con qualsiasi numero di alleli.

Il valore di PE dipende dalla frequenza degli alleli presenti nei marcatori utilizzati. Evidentemente, se $p = 1$, il gene è monomorfico, e $PE = 0$. Se $p \neq 1$, la probabilità di esclusione aumenta fino ad un valore che, in un sistema di due alleli codominanti, raggiunge il massimo quando $p = q = 0.5$. PE aumenta all'aumentare del numero di alleli. I valori massimi di PE in sistemi codominanti con k alleli con frequenze uguali ($1/k$), sono riportati in tabella 5.

Se PE è stimata in sistemi multilocus (ciascuno con probabilità di

Tabella 3 - Probabilità di esclusione per un locus con due alleli codominanti. np = combinazione allelica non possibile.

Genotipi madre/figlio	Frequenze alleliche	Genotipi paterni			Probabilità di esclusione
		AA	Aa	aa	
AA - AA	p^3	0	0	x	p^3q^2
Aa - AA	p^2q	x	0	0	p^2q
aa - AA	np	--	--	--	--
AA - Aa	p^2q	0	0	x	p^2q^3
Aa - Aa	$p^2q + pq^2$	0	0	0	--
aa - Aa	pq^2	x	0	0	p^3q^2
AA - aa	np	--	--	--	--
Aa - aa	pq^2	0	0	x	pq^4
aa - aa	q^3	x	0	0	p^2q^3

Tabella 4 - Probabilità di esclusione dei genotipi paterni per ogni coppia di genotipi madre-figlio per sistemi codominanti qualsiasi numero k di alleli.

Madre		Figlio		Pr esclusione padre	
Genotipo	Pr	Genotipo	Pr	Genotipo	Pr
$a_i a_i$	p_i^2	$a_i a_i$	p_i	$a_w a_x (w, x \# i)$	$(1-p_i)^2$
		$a_i a_j$	p_j	$a_w a_x (w, x \# j)$	$(1-p_j)^2$
$a_i a_j$	$2p_i p_j$	$a_i a_i$	$p_i/2$	$a_w a_x (w, x \# i)$	$(1-p_i)^2$
		$a_i a_j$	$p_j/2$	$a_w a_x (w, x \# j)$	$(1-p_j)^2$
		$a_i a_j$	$(p_i+p_j)/2$	$a_w a_x (w, x \# i, j)$	$(1-p_i - p_j)^2$
		$a_i a_k$	$p_k/2$	$a_w a_x (w, x \# k)$	$(1-p_k)^2$
		$a_i a_k$	$p_k/2$	$a_w a_x (w, x \# k)$	$(1-p_k)^2$
		$a_i a_k$	$p_k/2$	$a_w a_x (w, x \# k)$	$(1-p_k)^2$

esclusione media PE_i), il valore medio di esclusione è (regola del prodotto):

$$PE_m = 1 - (1 - PE_1) (1 - PE_2) \dots (1 - PE_k)$$

Questa formula è valida se i marcatori usati sono statisticamente indipendenti.

Probabilità di match

E' la probabilità che un genotipo corrisponda al genotipo di un individuo estratto a caso dalla popolazione di riferimento. Quando i genotipi ottenuti da due campioni sono identici: $G_c = G_s = G$, occorre verificare le probabilità che, date tutte le condizioni di contorno, essi appartengano allo stesso individuo. Quindi dobbiamo confrontare la probabilità di due ipotesi alternative:

- *H_p*: G_c e G_s appartengono allo stesso individuo
- *H_d*: G_c e G_s appartengono a due diversi individui

Tabella 5 - Massima probabilità di esclusione (non paternità) per sistemi genetici codominanti con k alleli.

k	Pr massima	
2	3/16	0.188
3	30/81	0.370
4	129/256	0.504
5	372/625	0.595
6	855/1296	0.660
7	1698/2401	0.707
8	3045/4096	0.743
9	5064/6561	0.772
10	7947/10000	0.795

La verosimiglianza relativa delle due ipotesi alternative è valutata tramite LR . Un data-base di frequenze alleliche della popolazione o della sottopopolazione di riferimento, può essere usato per stimare le frequenze dei genotipi. Il valore di LR dipende da una serie di assunzioni: i due individui sono o non sono imparentati; il campione usato per stimare le frequenze alleliche nella popolazione di riferimento è un campione casuale rappresentativo della popolazione; la stima delle frequenze genotipiche si ottiene tramite la regola del prodotto, che è valida se le frequenze alleliche sono reciprocamente indipendenti; la popolazione può essere o non essere strutturata. Se i due genotipi sono statisticamente indipendenti, allora $LR = 1 / \Pr(GdHd, I) = 1 / P$.

Stima della probabilità di match di un singolo locus biallelico. Se il campione è stato estratto da una popolazione in HWE, la probabilità di match ad un locus si calcola come:

11. $1 / P = p^2$: se il locus è omozigote
12. $1 / P = 2pq$: se il locus è eterozigote

Stima della probabilità di match di un genotipo multilocus. Se il campione è stato estratto da una popolazione in HWE, la probabilità di match di un genotipo multilocus si calcola tramite la regola del prodotto.

In alcuni casi i due genotipi non sono indipendenti a causa delle relazioni familiari, oppure a causa dell'inbreeding nella popolazione o nella sottopopolazione. Nel confronto fra i genotipi di due individui sono coinvolti quattro alleli per ogni locus (due alleli materni e due alleli paterni), che possono, in parte o totalmente, essere ibd. Assumendo che i genitori non siano inbred, è possibile prevedere le probabilità di ibd per coppie di individui imparentati. Da queste probabilità si possono ricavare le equazioni per calcolare le probabilità di match fra due campioni (G_1

Tabella 6 - Probabilità che due individui (1 e 2) imparentati abbiano lo stesso genotipo (G_1 o G_2) omozigote (a_1a_1) o eterozigote (a_1a_2).

Relazione di parentela	Genotipo $G_1 = a_1a_1$	Genotipo $G_2 = a_1a_2$
Genitore - figlio	p_i	$(p_i + p_j)/2$
Fratelli	$(1 + p_i)^2/4$	$(1 + p_i + p_j + 2p_i p_j)/4$
Nonno - nipote	$p_i(1 + p_i)/2$	$(p_i + p_j + 4p_i p_j)/4$
Fratellastri	$p_i(1 + p_i)/2$	$(p_i + p_j + 4p_i p_j)/4$
Zio - nipote	$p_i(1 + p_i)/2$	$(p_i + p_j + 4p_i p_j)/4$
Primi cugini	$p_i(1 + 3p_i)/4$	$(p_i + p_j + 12p_i p_j)/8$

e G_2) che hanno lo stesso genotipo omozigote ($a_i a_i$) o eterozigote ($a_i a_j$) nell'ipotesi in cui essi siano imparentati (Tab. 6).

Effetto dell'inbreeding nella popolazione.

Se la popolazione è di piccole dimensioni, oppure è composta da sottopopolazioni distinte e di piccole dimensioni, allora due individui non imparentati presi a caso sono in qualche misura inbred. Le probabilità di match nel caso in cui due individui abbiano lo stesso genotipo omozigote o eterozigote sono calcolate tramite le equazioni di Balding e Nichols (1994):

- genotipi omozigoti: $\Pr(G_1 = a_i a_i | G_2 = a_i a_i) = [2\theta + (1 - \theta)p_i][3\theta + (1 - \theta)p_i] / (1 - \theta)(1 - 2\theta)$
- genotipi eterozigoti: $\Pr(G_1 = a_i a_j | G_2 = a_i a_j) = 2[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j] / (1 - \theta)(1 - 2\theta)$

La quantità θ descrive il grado di inbreeding medio fra i membri della sottopopolazione relativamente alla popolazione totale. L'equazione assume HWE nella sottopopolazione, ma scostamenti dall'equilibrio nella popolazione totale. Per calcolare θ occorre disporre di data-base sulle frequenze alleliche nelle sottopopolazioni dei marcatori genetici utilizzati. Raramente queste informazioni sono disponibili. Il rapporto NRC 1996 raccomanda di usare, nelle popolazioni umane, valori di $\theta = 0.01 - 0.03$.

Gli effetti numerici della correzione θ sono di solito piccoli, a meno che le frequenze alleliche siano piccole e θ sia grande (Tab. 7). Spesso si utilizza la regola del prodotto anche se il test di indipendenza è significa-

Tabella 7 - Effetto dell'uso delle equazioni di Balding e Nichols (1994) sulla stima delle probabilità di match. Valori di LR per genotipi eterozigoti in una popolazione con frequenze alleliche p_i e struttura θ . Valori di LR per genotipi omozigoti.

	$\theta = 0$	$\theta = 0,001$	$\theta = 0,01$	$\theta = 0,03$
<i>G</i> eterozigote				
$p_i = 0,01$	5.000	4.152	1.295	346
$p_i = 0,05$	200	193	145	89
$p_i = 0,10$	50	49	43	43
<i>G</i> omozigote				
$p_i = 0,01$	10.000	6.439	863	157
$p_i = 0,05$	400	364	186	73
$p_i = 0,10$	100	96	67	37

tivo, cioè se l'ipotesi nulla di indipendenza è falsa, assumendo che l'effetto della dipendenza sulle stime delle frequenze genotipiche sia piccolo.

Probabilità di identità (*Pid*)

È la probabilità che due individui estratti a caso dalla stessa popolazione abbiano lo stesso genotipo multilocus. *Pid* corrisponde alla sommatoria del quadrato della probabilità di match per ognuno dei loci che compongono il genotipo multilocus.

La *Pid* attesa ad ogni locus viene calcolata usando le frequenze alleliche p stimate ad un campione della popolazione:

- $Pid = \sum p_i^4 + \sum \sum (2p_i p_j)^2$

Questa formula può essere facilmente modificata per calcolare la *Pid* attesa fra fratelli:

- $Pid_{sib} = 0,25 + (0,5 \sum p_i^2) + [0,5 (\sum p_i^2)^2] - (0,25 \sum p_i^4)$

Pid e Pid_{sib} possono essere stimate per qualsiasi numero di loci usando il software Prob-ID3 (G. Luikart).

Nei sistemi di DNA fingerprinting MLP le stime di probabilità di identità vengono calcolate diversamente. La probabilità che ogni allele di un DNA fingerprinting presente in un individuo *A* sia anche presente in un individuo *B* scelto a caso nella popolazione cui *A* e *B* appartengono, dipende dalla frequenza dell'allele nella popolazione. Se la frequenza dell'allele è q , allora la probabilità che un individuo scelto a caso dalla popolazione contenga quest'allele è $x = 2q - q^2$ (probabilità di match), formulazione che deriva dalla legge di Hardy-Weinberg. L'eterozigosità, cioè la proporzione degli individui che possiedono l'allele allo stato eterozigote è $h = (2q - 2q^2)/(2q - q^2) = 2(1 - q)/(2 - q)$. Per la stima di x si determina il DNA fingerprinting di una serie di campioni. Si identificano i frammenti dei profili genetici individuali, e si calcolano le proporzioni dei campioni che possiedono ciascuno dei frammenti. La media di queste proporzioni è la stima della probabilità media di match $x = 2q - q^2$, oppure $x = 2q$, se la frequenza q è sufficientemente piccola e q^2 è molto minore di q . Da questo valore si calcola la frequenza allelica q , da cui si ottiene la stima dell'eterozigosità h . Il numero medio di frammenti m di un DNA fingerprinting è dato dal numero totale di frammenti individuati diviso il numero di individui analizzati. La probabilità che un individuo scelto a caso abbia un DNA fingerprinting identico all'individuo analizzato è: x^m .

La probabilità di identità *Pid* non è equivalente alla probabilità di

match. Infatti P_{id} si riferisce alla comparazione di due campioni, ed è la probabilità che due individui scelti a caso abbiano lo stesso genotipo. Invece la probabilità di match si riferisce alla comparazione di un singolo individuo ad una serie di genotipi, ed è la probabilità che un determinato individuo sia identico ad una serie di genotipi.

La probabilità di discriminazione: $P_{dis} = 1 - P_{id}$ è la probabilità che due individui scelti a caso dalla stessa popolazione siano distinguibili usando la stessa serie di marcatori genetici. I valori massimi di P_{dis} sono ottenuti quando le frequenze alleliche ad ogni locus sono uguali ($1/k$). In questo caso, la massima $P_{dis} = 1 - (2k - 1)/k^2$.

Test di paternità

Il teorema di Bayes è molto utile nei test di paternità o di parentela. Un'ipotesi di paternità rappresenta un evento incerto (H). Sono disponibili alcune informazioni (I) che possono condizionare l'incertezza e altre informazioni (E) che costituiscono l'evidenza. In genetica forense l'evidenza è costituita dal DNA fingerprinting. Vogliamo valutare come l'evidenza E possa contribuire a stimare la probabilità di H . La probabilità che un maschio incluso, a seguito di analisi di DNA fingerprinting, sia il padre biologico è la probabilità di paternità (assumendo che la madre ed i suoi figli siano genotipizzati). La probabilità è condizionale e si applica solo a quei maschi che non sono stati esclusi come possibili padri in seguito alle analisi genetiche. Un individuo che non viene escluso, viene automaticamente incluso e considerato come il possibile genitore.

Il teorema di Bayes consente di valutare le due ipotesi alternative:

- H_p = probabilità che il padre presunto (non-escluso) **sia** il padre biologico, dati H ed I
- H_d = probabilità che il padre presunto (non-escluso) **non sia** il padre biologico, dati H ed I

Il teorema di Bayes si può scrivere nelle forma seguente (pronostici):

$$\bullet \Pr(H_p|E, I) / \Pr(H_d|E, I) = \Pr(E|H_p, I) / \Pr(E|H_d, I) \times \Pr(H_p|I) / \Pr(H_d|I)$$

Per calcolare la probabilità di paternità occorre conoscere le due probabilità condizionali:

- $\Pr(E|H_p, I)$: la Pr che il padre presunto (non-escluso) sia il padre biologico, assumendo che esso sia il padre biologico = $\Pr(\text{non escluso}|\text{padre biologico})$
- $\Pr(E|H_d, I)$: la Pr che il padre presunto sia un'individuo qualsiasi

(random rispetto al suo genotipo) che per caso non è stato escluso: = $\Pr(\text{non escluso}|\text{random})$

e le due probabilità a-priori:

- $\Pr(Hp|I)$: la Pr che un altro individuo possa essere il padre biologico, indipendentemente dall'evidenza genetica = $\Pr(\text{padre biologico})$
- $\Pr(Hd|I)$: la probabilità inversa: $\Pr(\text{random}) = 1 - \Pr(\text{padre biologico})$, cioè la probabilità che il padre presunto sia un'estraneo scelto a caso, indipendentemente dal suo genotipo

Il teorema di Bayes è equivalente alla forma seguente:

- pronostico a-posteriori = $LR \times$ pronostico a-priori.

Assumiamo che il padre presunto (non-escluso) sia il padre biologico. Allora la $\Pr(\text{non escluso}|\text{padre biologico})$ si calcola dal confronto fra gli alleli del padre presunto e gli alleli paterni che sono presenti nei figli. Un figlio deve avere un allele identico ad almeno un allele materno, l'altro allele deve derivare dal padre. In un sistema biallelico codominante, se il padre presunto è omozigote ha $\Pr = 1$ di aver trasmesso il suo allele al figlio; se è eterozigote ha $\Pr = 0.5$. Il prodotto delle Pr ad ognuno dei loci che compongono il profilo multilocus corrisponde alla Pr totale che

Tabella 8 - Probabilità del genotipo del figlio, assumendo che le due ipotesi alternative: $Hp = \Pr(\text{non escluso}|\text{padre biologico}) = (G_C|G_M, G_{AF}, H_p)$, ed $Hd = \Pr(\text{non escluso}|\text{random}) = \Pr(G_C|G_M, G_{AF}, H_d)$, siano vere. Si indicano, come esempio, i valori di $LR = Hp / Hd = \Pr(G_C|G_M, G_{AF}, H_p) / \Pr(G_C|G_M, G_{AF}, H_d)$, per frequenze alleliche $p_i = p_j = 0,1$

G_C	G_M	G_{AF}	H_p	H_d	LR	$LR; p_i = p_j = 0,1$		
$a_i a_i$	$a_i a_i$	$a_i a_i$	1	p_i	$1/p_i$	10		
		$a_i a_j$	0,5	p_i	$1/2p_i$	5		
		$a_i a_k$	0	p_i	0	0		
	$a_i a_j$	$a_i a_j$	$a_i a_i$	0,5	$p_i/2$	$1/p_i$	10	
			$a_i a_j$	0,25	$p_i/2$	$1/2p_i$	5	
			$a_i a_k$	0	$p_i/2$	0	0	
		$a_i a_k$	$a_i a_k$	$a_j a_j$	1	p_i	$1/p_j$	10
				$a_j a_k$	0,5	p_i	$1/2p_j$	5
				$a_k a_l$	0	p_i	0	0
$a_j a_j$	$a_i a_j$	$a_i a_i$	0,5	$(p_i + p_j)/2$	$1/(p_i + p_j)$	5		
		$a_i a_j$	0,5	$(p_i + p_j)/2$	$1/(p_i + p_j)$	5		
		$a_j a_k$	0,25	$(p_i + p_j)/2$	$1/2(p_i + p_j)$	2,5		
	$a_j a_k$	$a_j a_k$	$a_k a_l$	0	$(p_i + p_j)/2$	0	0	
			$a_j a_j$	0,5	$p_j/2$	$1/p_j$	10	
			$a_j a_l$	0,25	$p_j/2$	$1/2p_j$	5	
		$a_k a_l$	$a_k a_l$	$a_k a_l$	0	$p_j/2$	0	0

il padre presunto abbia trasmesso gli alleli non-materni ai figli = $\Pr(\text{non escluso}|\text{padre biologico})$. Questa è la probabilità che il padre presunto abbia prodotto il genotipo paterno del figlio (Tab. 8).

Assumiamo che il padre presunto non sia il padre biologico, ma un individuo qualsiasi che non è stato escluso per caso, cioè che ha per caso un genotipo compatibile col padre presunto. In alcuni casi è possibile identificare padri presunti alternativi, cioè più individui che potrebbero essere i padri biologici. Così le probabilità di individui alternativi, oppure del padre presunto e di un individuo qualsiasi preso a caso, possono essere comparate. Il significato del confronto è di accertare quanto sia più (o meno) probabile che il padre presunto sia il padre biologico. La stima di probabilità di paternità non costituisce un valore assoluto, è relativa la confronto con padri presunti alternativi, oppure con un'individuo preso a caso dalla popolazione di riferimento. Questa probabilità si ottiene tramite il calcolo del prodotto delle frequenze degli alleli paterni osservati nel figlio. Le frequenze alleliche sono quelle della popolazione di riferimento.

Il rapporto fra queste due probabilità condizionali è l'indice di paternità:

$$r = \Pr(\text{non escluso}|\text{padre biologico}) / \Pr(\text{non escluso}|\text{random})$$

Quanto più r è grande, tanto più è verosimile che il padre presunto sia il padre biologico. Se esistono padri presunti alternativi, r è calcolato per ogni coppia di padri alternativi ed il valore più grande indica qual'è il padre biologico più verosimile. Non è necessario considerare un individuo random. Per valori di $r > 40$ la probabilità di paternità diventa praticamente 1, quando esistono uguali probabilità a-priori. Assumendo che la probabilità a-priori del padre presunto di essere il padre biologico sia 0.75 (e quindi la probabilità di un individuo qualsiasi sia 0.25), la probabilità di paternità tende ad uno anche usando pochi loci diagnostici.

Le due probabilità a-priori devono essere determinate indipendentemente dall'evidenza genetica. Spesso è difficile stabilire queste probabilità a-priori. Se non è possibile fare assunzioni a-priori allora si assume che: $\Pr(\text{padre biologico}) = \Pr(\text{random}) = 0,5$. Disponendo delle probabilità a-priori è possibile calcolare la probabilità di paternità a-posteriori tramite il teorema di Bayes:

$$\Pr(\text{padre biologico}|\text{non escluso}) = [\Pr(\text{non escluso}|\text{padre biologico})\Pr(\text{padre biologico})] / \Pr(\text{random})$$

Questa formulazione può essere trasformata in:

$$\Pr(\text{padre biologico}|\text{non escluso}) = 1 / \{1 + [\Pr(\text{random}) / \Pr(\text{padre biologico})] \times [\Pr(\text{non escluso}|\text{random}) / \Pr(\text{non escluso}|\text{padre biologico})]\}$$

Se le assunzioni di Pr a-priori sono equivalenti, la formula diventa:

$$\bullet \Pr(\text{padre biologico} | \text{non escluso}) = 1 / \{1 + [\Pr(\text{non escluso} | \text{random}) / \Pr(\text{non escluso} | \text{padre biologico})]\} = 1 / (1 + 1 / r)$$

L'indice di paternità può essere formulato anche come:

$$\bullet LR = \Pr(G_C | G_M, G_{AF}, H_p) / \Pr(G_C | G_M, G_{AF}, H_d)$$

Con:

G_C = genotipo del figlio C

G_M = genotipo della madre biologica M di C

G_{AF} = genotipo del padre presunto AF di C

H_p = il padre presunto AF è il padre biologico di C

H_d = qualcun altro (non imparentato con AF) è il padre (alternativo) di C
 AF e M non sono imparentati.

Se l'ipotesi H_d è che il padre presunto alternativo sia imparentato con il padre indiziato, allora le probabilità del denominatore di LR cambiano. Occorre introdurre stime di probabilità di ibd e di coancestry (q) fra gli alleli e fra i genotipi dei due possibili padri. Se si assume che il padre alternativo non sia inbred, allora $ibd = F = 0$. I valori di LR dovranno solo tener conto dell'effetto di coancestry stimato da q .

E' possibile stimare i valori di LR quando non esiste il genotipo del padre presunto, ma esiste il genotipo di un suo parente R . In questo caso le ipotesi diventano:

- H_p = il padre di C è parente di R
- H_d = il padre di C non è parente di R

Si dimostra che (in assenza di inbreeding):

$$\bullet LR = (1 - 2q_{AR}) + 2q_{AR}r$$

Dove:

r = indice di paternità; q_{AR} = coefficiente di coancestry per il padre presunto ed il parente di cui esiste in genotipo, che vale: 1/4 per fratelli e per padre e figlio; 1/8 per fratellastri o per zio e nipote; 1/16 per i primi cugini.

Popolazioni strutturate. In popolazioni strutturate non si può assumere che gli individui siano non imparentati, ma occorre assumere che esista in certo livello di parentela fra di loro. Quindi la madre, il padre presunto ed il padre alternativo sono in qualche modo geneticamente interrelati, anche se non appartengono alla stessa famiglia. Se sono note le frequenze alleliche per la sottopopolazione da cui deriva il nucleo fami-

liare da testare, allora i valori di LR si possono stabilire esattamente come sopra (Tabella 8). Ma se conosciamo solo i valori di frequenze alleliche nella popolazione totale, allora è necessario valutare la divergenza genetica fra le sottopopolazioni. Nella formulazione per calcolare $LR = \Pr(G_C | G_M, G_{AP}, H_p) / \Pr(G_C | G_M, G_{AP}, H_d)$, il numeratore non cambia, ma nella stima del denominatore non è possibile assumere che i genotipi materni e paterni siano indipendenti.

CASI DI STUDIO

Presso il Laboratorio di genetica dell'INFS, nell'ambito di un accordo quadro con il Ministero dell'Ambiente, Divisione II, Direzione Conservazione della Natura, sono in corso analisi di genetica forense in applicazione della CITES. Il grafico in figura 44 mostra il numero di analisi genetiche richieste ed effettuate per la certificazione CITES di nascita in cattività, in specie di uccelli.

I test di parentela negli uccelli vengono effettuati mediante due metodologie.

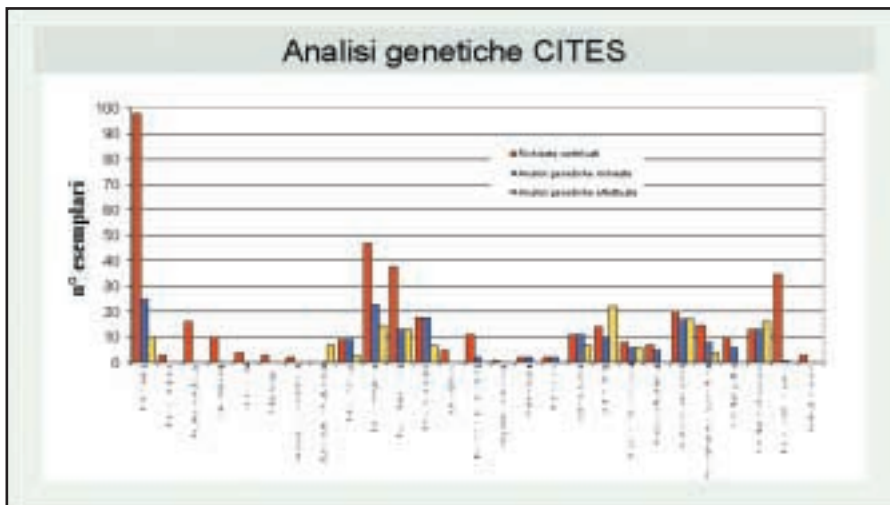


Figura 44 - Esempio di specie e numero di gruppi familiari sottoposti al controllo delle nascite in cattività tramite test genetici di parentela presso il Laboratorio di Genetica dell'Istituto Nazionale per la Fauna Selvatica, nell'ambito dell'attività CITES.

Test di parentela effettuati tramite DNA fingerprinting MLP

In alcune specie di Psittaciformi e di Strigiformi non sono disponibili microsatelliti, per cui è necessario effettuare le analisi di parentela tramite DNA fingerprinting MLP. In figura 45 si riportano gli esempi di un caso di inclusione, in cui i due genitori presunti sono stati identificati come probabili genitori biologici dei due figli, e di un caso di esclusione, in cui uno dei due genitori presunti è stato escluso come possibile genitore.

La variabilità genetica ai sistemi di DNA fingerprinting viene identificata tramite digestione del DNA dei campioni con gli enzimi di restrizione AluI e HaeIII, a cui segue l'ibridazione con le sonde multilocus di Jeffreys 33.15 e 33.6. Queste sonde individuano da 20 a 30 frammenti per individuo, che sono polimorfici in tutte le specie di Psittaciformi ed altri uccelli, in un range di pesi molecolari che vanno da 3.5 a 20 kb. Solo circa l'1-2% dei frammenti sono coidentificati da entrambe le sonde, ed i frammenti sono di solito in LE. L'eterozigosi ed i valori di probabilità di identità sono stati stimati in tre specie di Psittaciformi in cui è stato

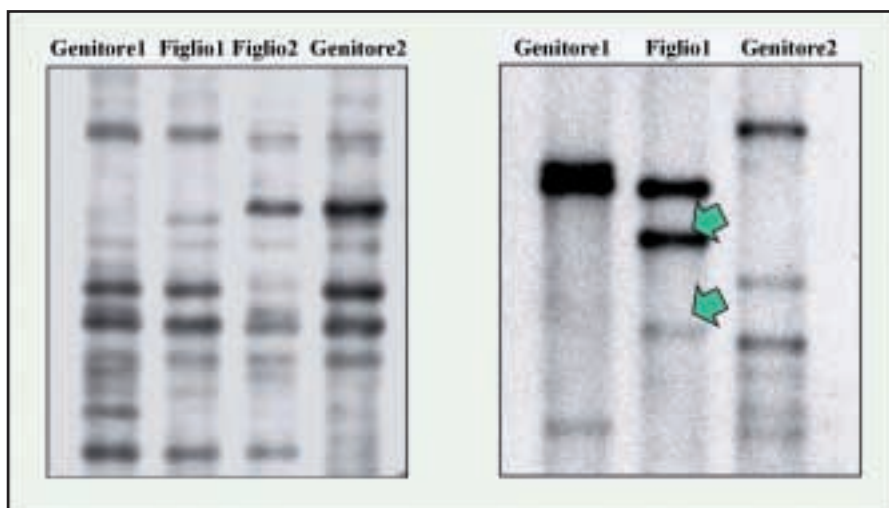


Figura 45 - Test di parentela in due famiglie di psittaciformi, effettuati tramite analisi di DNA fingerprinting multilocus (MLP). Nel primo caso (riquadro a sinistra), tutti i frammenti di DNA (alleli) presenti nei due figli sono presenti anche in uno o nell'altro dei due presunti genitori. I genotipi dei figli sono compatibili con i genotipi dei presunti genitori. Non è pertanto possibile escludere che i genitori presunti siano i genitori biologici dei figli. Nel secondo caso (riquadro a destra), il figlio n. 1 presenta due frammenti su tre che non sono presenti nei genitori. E' possibile, pertanto, escludere che questo individuo sia il figlio naturale dei due genitori putativi.

Tabella 9 - Variabilità in DNA fingerprinting determinati in 30 campioni.

Campione/ Frammento	1	2	3	4	5	6	7	8	9	10	11	12	13	Frequenza
1							1			1				0,15
2			1		1	1				1				0,31
3							1	1						0,15
4	1	1										1		0,31
5			1						1					0,08
6						1								0,08
7													1	0,08
8	1	1	1		1	1	1	1	1	1	1	1	1	0,92
9				1					1					0,15
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12				1				1	1		1			0,31
13													1	0,08
14													1	0,08
15							1				1		1	0,23
16							1				1		1	0,23
17							1	1			1		1	0,31
18												1	1	0,15
19				1	1	1				1				0,31
20									1					0,08
21	1	1		1							1	1		0,38
22	1	1	1		1	1								0,38
23			1	1		1	1	1	1	1	1	1		0,77
24	1	1	1	1			1	1			1		1	0,61
25								1	1					0,15
26			1											0,08
27			1	1		1	1	1		1	1	1		0,61
28					1							1		0,15
29	1	1						1			1	1	1	0,46
30	1		1										1	0,23

possibile ottenere almeno 10 individui (riprodotti in cattività) presumibilmente non imparentati, usando il metodo di Jeffreys. I risultati sono riportati nelle tabelle 9 e 10.

I valori necessari per calcolare la probabilità di identità sono:

- 1) il numero medio di frammenti per individuo è $m = \text{no. di frammenti totali} / \text{no. di individui}$ (che nell'esempio preso in considerazione è $m = 128/13 = 9,85$);
- 2) la probabilità di identità media $x = \Sigma \text{ frequenza di ogni frammento} /$

Tabella 10 - Probabilità di identità in specie di psittaciformi.

Enzima Specie	A1uI			HaeIII		
	<i>Amazona viridiginalis</i>	<i>Cacatua moluccensis</i>	<i>Ara macao</i>	<i>Amazona viridiginalis</i>	<i>Cacatua moluccensis</i>	<i>Ara macao</i>
M	9,84	10,25	9,75	11,09	10,10	8,33
X medio	0,33	0,23	0,25	0,44	0,37	0,33
Pid	$1,72 \times 10^{-5}$	$2,60 \times 10^{-7}$	$1,35 \times 10^{-6}$	$1,22 \times 10^{-4}$	$4,86 \times 10^{-5}$	$1,06 \times 10^{-4}$
Q medio	0,18	0,12	0,13	0,25	0,21	0,18
H medio	0,90	0,93	0,93	0,85	0,88	0,90

no. di frammenti (nell'esempio è $x = 9,85 / 30 = 0,33$).

Nell'esempio esaminato, la probabilità che due individui abbiano casualmente lo stesso profilo genetico è $x^m = 0,33^{9,85} = 1,72 \times 10^{-5}$.

La frequenza allelica media può essere calcolata risolvendo l'equazione $(2q - q^2) = 0,33$ quindi $q = 0,18$. L'eterozigosità osservata è $2(1 - q) / (2 - q) = 2(1 - 0,18) / (2 - 0,18) = 0,90$.

Test di parentela effettuati tramite microsatelliti

In altre specie di Psittaciformi e di Falconiformi esistono microsatelliti specifici che sono utilizzati per i test di paternità. Un esempio di elettroferogramma per l'analisi di microsatelliti in gruppi familiari di Psittaciformi e Falconiformi è presentato in figura 46.

Identificazione di sottospecie di scimpanzé mediante analisi del mtDNA.

Le sequenze nucleotidiche della regione di controllo del DNA mitocondriale consentono di identificare la sottospecie (materna) di scimpanzé.

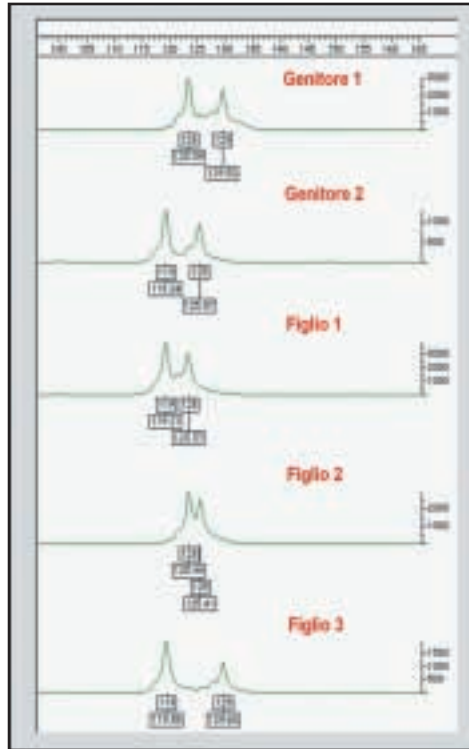


Figura 46 - Test di parentela in una famiglia di tigrì, effettuato tramite analisi automatica di un locus microsatellite (STR). I pesi molecolari degli alleli presenti nei due genitori (entrambi eterozigoti) sono determinati automaticamente. I figli presunti hanno genotipi (tutti eterozigoti) che contengono diverse combinazioni degli alleli dei due genitori. I genotipi dei figli sono compatibili con i genotipi dei presunti genitori. Non è pertanto possibile escludere che i genitori presunti siano i genitori biologici dei figli. Usando da 4 a 6 diversi loci microsatellite le probabilità di identità sono inferiori a 1/10.000, quindi è possibile concludere che è improbabile che i tre figli non derivino i loro genotipi dai due genitori putativi.

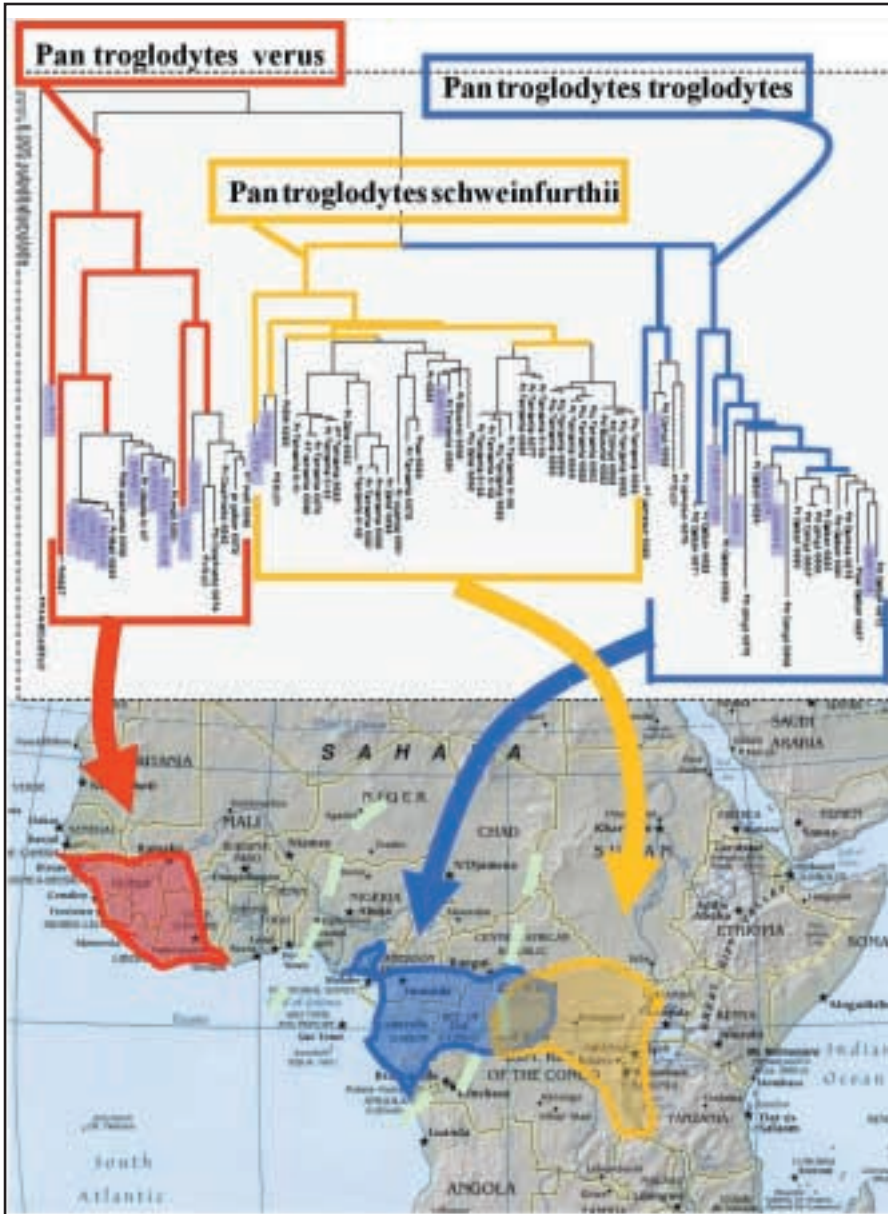


Figura 47 - Identificazione delle sottospecie di schimpanzè (*Pan troglodytes*) tramite analisi delle sequenze della regione di controllo del mtDNA. La distribuzione delle tre sottospecie è indicata con tre colori diversi nella mappa. A ciascuna delle tre sottospecie corrisponde un distinto ramo dell'albero filogenetico.

Le sequenze ottenute dagli esemplari da identificare sono allineate alle sequenze di riferimento. L'allineamento viene utilizzato per produrre un albero filogenetico che consente di identificare le sottospecie (Fig. 47).

EXECUTIVE SUMMARY

Randi E., C. Tabarroni e S. Rimondi (eds.), 2002 - *Forensic genetics and the Washington Convention - CITES*. Quad. Cons. Natura, 12, Min. Ambiente - Ist. Naz. Fauna Selvatica.

CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora), is an international agreement which aims at regulating the trade of plants and animals. It is based on the principle that control over the sustainable trade of fauna and flora and their products, constitutes a conservation measure of the wild populations, above all if the concept of sustainable use of living species, forms the basis of international and national legislation. In fact, the correct application of the CITES means that the dynamics of threatened species and populations subject to trade is constantly controlled. CITES works by authoring the issue of import and export permits of living specimens and their products that are among the protected species listed in Appendices I and II. The species in Appendix I are afforded total protection, and trade in specimens of these species is only permitted under exceptional circumstances. Trade of species listed in Appendix II is possible though must be closely controlled. CITES also regulates the detention and trade of fauna and flora reproduced in captivity and then possibly used for travelling collections or exhibitions. In these cases, CITES permits are issued only when there is proof that these specimens were born and bred in captivity. The EC Commission Regulation No. 1808/2001, regarding the protection of wild fauna and flora through trade regulations establishes that the Management Authorities of the State can avail themselves of genetic testing to determine the origin and the degree of kinship of plants and animals detained and reproduced in captivity. As a consequence of these norms, the Management Authority can issue export permits for commercial purposes of specimens listed in Appendix I reproduced in captivity, only after certification that the specimens in question were actually born in captivity.

Forensic genetics is going through a period of rapid progress thanks to the development of DNA molecular testing methodologies that have reached levels of precision, repeatability and reliability that were unthinkable until recently. The concept of DNA fingerprinting has rapidly become part of everyday speech. Molecular methodologies have an elevated capacity of identification (every individual, except for identical twins, has a unique genetic profile, that differs from any other individual). The results of laboratory tests can be interpreted in the context of population genetics and the theory of probability. In this manner the results of laboratory tests can be expressed in a quantitative manner (probabilistically) and evaluated through statistical analysis. The principal aim of forensic genetic testing is to verify the hypothesis that a specific DNA fingerprinting is univocally associated to a particular individual, or that the DNA fingerprinting of an offspring is derived from the DNA

fingerprinting of the two putative parents. The DNA testing methods permit the identification of every individual present in a population and the reconstruction of the degree of relationship within a family unit. The results of DNA tests provide information that can be used as evidence during legal proceedings in law courts. Forensic genetic procedures must guarantee high quality results, that must be evaluated accurately and be comprehensible also to those who are not geneticists by profession. Forensic genetic testing is used to provide the competent authorities with objective information that can assist them in making decisions and resolving legal disputes.

The methods used in molecular testing which allow the reconstruction of DNA fingerprinting are based on observations of the presence of very complex and variable DNA segment arrangements within genomes that are associated exclusively to each individual. The structure of DNA fingerprinting is caused by genetic mutations of the genes that are, almost always, well identified. The variability of DNA fingerprinting is rigorously analysed using models of population genetics and statistic procedures. The use of molecular genetics in forensic science is based on strong biological and statistical data. DNA fingerprinting is widely used in forensic genetics as well as in criminology, and is applied in decisions regarding paternity, identification of animal and plant species and individuals, poaching and trade of living specimens and their products. DNA fingerprinting testing can considerably reduce the level of subjectivity that is inherent in all identification procedures, as long as it is carried out and evaluated correctly. It is opportune to limit the definition of DNA fingerprinting to those methods of molecular testing that permit the identification of samples. These methods include: "DNA fingerprinting" recognition of typical multi-loci, achieved by means of multi-locus probes; multiple single locus (each one consisting of a variable number of tandem repeats); "DNA fingerprinting" recognition, attained by means of specific single locus probes; PCR analysis of micro-satellite loci (short tandem repeats). Independently from which method is used, the pattern of DNA segments identified in each sample results in an individual genetic arrangement sample-specific.

At the INFS (National Institute for Wild Fauna) Genetic Laboratory, in concordance with the Ministry of the Environment, Nature Conservation Department, forensic genetic testing is currently underway in application of the CITES. The testing that is being carried out is principally for CITES certification of species born in captivity, and regard numerous bird and mammal species.

BIBLIOGRAFIA

- BALDING D. J. e R. A. NICHOLS, 1994 - *DNA profile match probability calculation*. *Forensic Sci. Int.*, 64: 125-140.
- DE KLEMM C., 1993 - *Guidelines for legislation to implement CITES*. IUCN, Gland, Switzerland.
- EVETT I. W. e B. S. WEIR, 1998 - *Interpreting DNA evidence*. Sinauer, Sunderland, MA.
- NRC - NATIONAL RESEARCH COUNCIL, 1966 - *The Evaluation of Forensic DNA Evidence*. <http://bob.nap.edu/html/DNA/>
- HEMLEY G. (ed.), 1994 - *International wildlife trade*. Island Press, Washington D.C.
- INMAN K. e N. RUDIN, 1997 - *Forensic DNA analysis*. CRC Press, New York.
- JEFFREYS A. J., V. WILSON e S. L. THEIN, 1985 - *Hypervariable "minisatellite" regions in human DNA*. *Nature*, 314: 67-73.
- KIRBY L. T., 1990 - *DNA fingerprinting*. Stockton, New York.
- MEREU U., 1995 - *Commercio e tutela di animali e piante*. EdAs, Frascati, Roma.
- PRITCHARD J. K., M. STEPHENS e P. J. DONNELLY, 2000 - *Inference of population structure using multilocus genotype data*. *Genetics*, 155: 945-959.
- SAMBROOK J., E. F. FRITSCH e T. MANIATIS, 1996 - *Molecular cloning*. Cold Spring Harbor Laboratory Press.
- SOUTHERN E. M., 1975 - *Detection of specific sequences among DNA fragments separated by gel electrophoresis*. *Journal of Molecular Biology*, 98: 503-517.

Finito di stampare nel mese di maggio 2002
dalla Tipolitografia F.G. Savignano s/Panaro - Modena

