UNIVERSITÀ DEGLI STUDI DI ROMA "TOR VERGATA"

FACOLTA' DI MEDICINA E CHIRURGIA

DOTTORATO DI RICERCA IN

SCIENZE FORENSI

XXI CICLO

POLIMORFISMI DEL DNA MITOCONDRIALE: IMPLICAZIONI ANALITICHE E FORENSI

Dr.ssa Maria Rosa Barilaro

A.A. 2008/2009

Docente Guida/Tutor: Prof. Saverio Potenza

Coordinatore: Prof. Giovanni Arcudi

I N D I C E

Capitolo 1 – Il DNA Mitocondriale Umano

1.1 Cenni sulla biologia e la fisiologia del mitocondrio……1

1.1.1 Struttura ……..…… 1
1.1.2 La catena respiratoria……... 3
1.1.3 L'apoptosi……..... 6

1.2 Il DNA mitocondriale ……..…. 8

1.3 L'ereditarietà e la variabilità ……15

1.3.1 Ereditarietà materna…… 15
1.3.2 Tasso evolutivo e mutazioni…….....17
1.3.3 Mitocondriopatie derivanti da mutazione del DNA mitocondriale umano…… 23
1.3.3.1 Difetti nella sintesi proteica ……..……. 26
1.3.3.2 Difetti nei geni codificanti proteine ……... 27
1.3.4 Mitocondriopatie derivanti da mutazioni del DNA nucleare …… 28

1.4 Gli studi evoluzionistici ……...29

1.4.1 L'orologio molecolare ..…….…… 29
1.4.2 La variabilità del DNA mitocondriale applicata agli studi evoluzionistici di popolazioni umane…… 30

1.5 Gli aplogruppi mitocondriali e la storia dell'uomo…….. 37

1.5.1 Gli aplogruppi principali……. 39
1.5.2 Aplogruppi in Europa e in Italia……... 42
1.5.2.1 Geografia dell'Europa e dell'Italia …… 42

1.5.2.2 Storia archeologica dell'Europa …… 42
1.5.2.3 Storia archeologica dell'Italia ……... 45
1.5.2.4 Genetica dei polimorfismi classici e flussi migratori …… 47
1.5.2.5 Il DNA mitocondriale europeo ……. 48
1.5.2.6 Il DNA mitocondriale della popolazione italiana ……. 53

1.6 L'mtDNA nel forense: utilità e vantaggi……... 54

1.6.1 L'analisi del DNA mitocondriale nel forense……. 56
1.6.2 La distribuzione dei Caucasici in base ai dati del DNA mitocondriale…….. 57
1.6.3 La distribuzione dei Caucasici in base ai dati del DNA mitocondriale…….. 58

Capitolo 2 – Scopo della tesi

2.1 Sequenziamento totale del DNA mitocondriale……..…….. 59

2.2 Proposizione ragionata di SNPs utili per la discriminazione forense…..61

Capitolo 3 – Materiali e Metodi

3.1 Il campione …….. 63

3.2 L'estrazione del DNA totale …….. 64

3.3 Quantificazione del DNA ……... 64

3.4 L'amplificazione ……..... 65

3.4.1 I parametri …… 73
3.4.2 Amplificazione della zona 8FR ……..……. 73
3.4.3 Purificazione dei prodotti PCR ……..…….. 74
3.4.4 La verifica su gel ……..…….75

3.5 Il sequenziamento ……..……. 76

3.5.1 Le reazioni di sequenza ……..……... 76

3.5.2 I parametri …..... 77

3.5.3 Purificazione delle reazioni di sequenza ……..... 78

3.5.4 L'elettroforesi di sequenziamento…….. 78

3.6 Analisi informatica ……..…….…....79

Capitolo 4 – Risultati e discussione

4.1 Analisi delle sequenze …… 82

4.2 Eteroplasmie di lunghezza …… 87

4.2.1 Polimorfismi noti in letteratura ……..…….……. 88

4.2.1.1 Polimorfismo poliC-303 …… 88

4.2.1.2 Polimorfismo poliC-16184 …… 91

4.2.1.3 Polimorfismo poliC-568 …… 95

4.2.2 Polimorfismi nuovi …….…….….. 96

4.3 Inserzioni e delezioni ……..… 97

4.4 Eteroplasmie ……. 100

4.5 Variazioni polimorfiche a singolo nucleotide …… 102

4.5.1 Polimorfismi nei geni dei tRNA ……... 104

4.5.2 Polimorfismi nei geni degli rRNA …… 108

4.5.3 Polimorfismi nei geni codificanti le subunità dei complessi enzimatici coinvolti nella fosforilazione ossidativa……..……..... 109

4.5.3.1 Polimorfismi nei geni codificanti le subunità della citocromo *c* ossidasi …….. 109

4.5.3.2 Polimorfismi nei geni codificanti le subunità 6 e 8 dell'ATP sintasi …….. 110

4.5.3.3 Polimorfismi che coinvolgono il gene codificante il citocromo b ….. 112

4.5.3.4 Polimorfismi nei geni che codificano per le subunità della NADH deidrogenasi ……. 113
4.5.4 Polimorfismi nelle regioni non codificanti……... 117
4.5.5 Analisi dei polimorfismi……... 119

4.6 Classificazione dei campioni analizzati in aplogruppi…….. 123

4.6.1 Individuazione di SNPs utili: quali risvolti?……..…... 144
4.6.1.1 Un caso problematico: il campione GEA21…….…….. 152
4.6.2 Quale database mitocondriale?..... 154

Capitolo 5 – Conclusioni e prospettive

5.1 Conclusioni derivanti dal presente lavoro…….. 157

5.2 …Prospettive di maggiore portata…….. 159

BIBLIOGRAFIA……..…166

Capitolo 1

IL DNA MITOCONDRIALE UMANO

1.1 Cenni sulla biologia e la fisiologia del mitocondrio

1.1.1 Struttura

I mitocondri sono organelli cellulari, di origine endosimbiontica, della dimensione di circa 1 μm (fig. 1.1 a). Sono presenti sia nelle cellule animali che in quelle vegetali in quantità variabile, da varie centinaia a un migliaio. Poiché i mitocondri hanno il ruolo di centrali energetiche cellulari in genere i tessuti metabolicamente attivi sono quelli che ne contengono il maggior numero. In media i mitocondri occupano circa il 30 % del volume citoplasmatico.

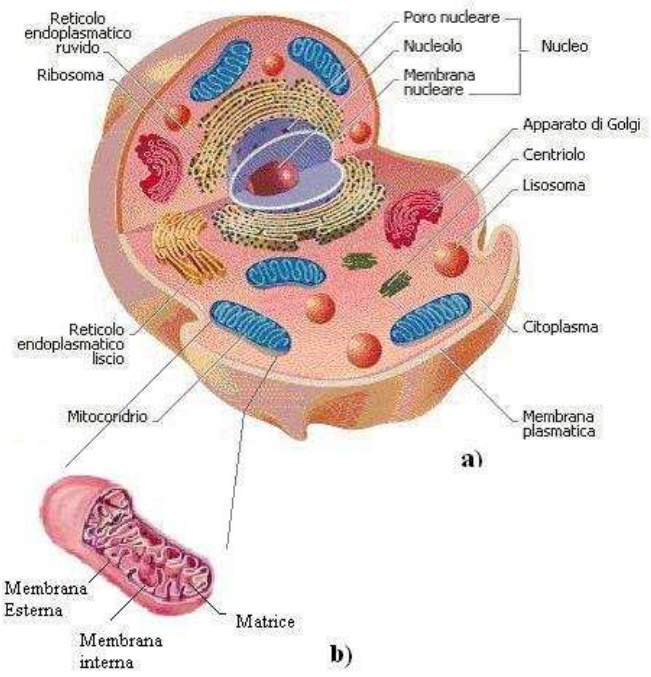


Fig. 1.1 a) Cellula eucariotica b) Dettaglio del mitocondrio. Sono mostrate la membrana esterna (permeabile a piccole molecole e ioni), la membrana interna formante le creste (impermeabile alla maggior parte delle piccole molecole e degli ioni, compresi i protoni, e contenente i complessi della catena respiratoria, l'ATP – ADP traslocasi e altri trasportatori di membrana) e la matrice mitocondriale (contenente il complesso della piruvato deidrogenasi, gli enzimi dell'acido citrico, gli enzimi della β-ossidazione degli acidi grassi, DNA, ribosomi, ATP,ADP, ioni ed enzimi).

I mitocondri possiedono una membrana esterna liscia ed una interna

dotata di creste. Nello spazio interno è presente una soluzione acquosa, la matrice che, al pari dello spazio intramebrana è molto ricca di enzimi ossidativi (Fig. 1.1b).

Dallo studio della struttura, funzione, morfologia e proprietà biochimiche e fisiologiche dei mitocondri e del mtDNA nel 1971 fu suggerita un'ipotesi sulla loro origine, che coinvolge la teoria endosimbiontica. Questa ipotesi prevedeva che le cellule eucariotiche si fossero originate in due fasi: prima avrebbe avuto origine il nucleo, a partire da un archeobatterio, poi sarebbe cominciato il rapporto simbiotico con il progenitore batterico dell'attuale mitocondrio. Secondo questa teoria gli archeozoi privi di mitocondri sarebbero alla base dell'albero degli eucarioti, in quanto precedenti alla formazione degli organelli. Più di recente si è affacciata sulla scena una nuova teoria, detta ipotesi dell'idrogeno (fig. 1.2), che a differenza di quella originaria immagina la creazione simultanea del nucleo e degli organelli, tramite fusione di un archeobatterio metanogeno (ospite), richiedente idrogeno, e un α-protobatterio (simbionte), fornitore dell'idrogeno.

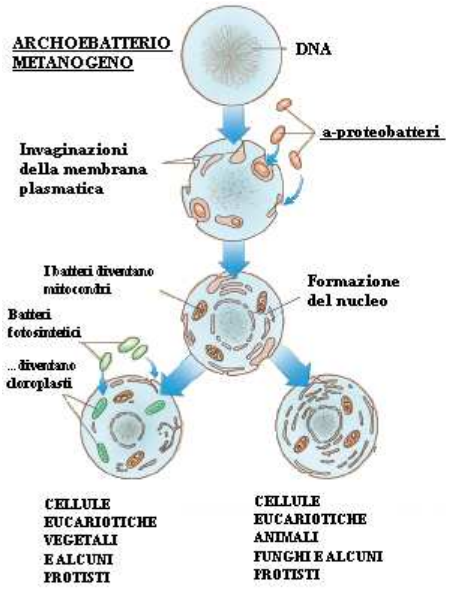


Fig. 1.2 Schema illustrante l'ipotesi dell'idrogeno. I proteobatteri vengono fagocitati in invaginazioni della membrana plasmatica degli archeobatteri metanogeni; in seguito parte del DNA dei proteobatteri è trasferito a quello della cellula ospite, rendendo così i primi totalmente dipendenti dai secondi.

Questa ipotesi è avallata da numerose osservazioni: a) il genoma nucleare eucariotico risulta costituito in parte da genoma archeobatterico e in parte da genoma eubatterico; b) nel genoma nucleare degli archeozoi sono presenti geni simili a quelli mitocondriali; c) la somiglianza che sussiste tra le proteine codificate da idrogenosomi e mitocondri. In ogni caso, secondo entrambe le teorie, l'endosimbionte putativo avrebbe trasmesso, col passare del tempo, parte del suo genoma al nucleo dell'ospite, conservando solo una piccola parte di geni tuttora compartimentati negli organelli. Ciò spiega anche le differenti dimensioni di questi genomi presenti nelle varie specie animali e vegetali. E' da sottolineare anche che, ciononostante, il gruppo di geni mantenuti negli organelli sono più o meno costanti (Gray et al, 1999).

1.1.2 La catena respiratoria

I mitocondri sono la sede delle reazioni chimiche che portano alla produzione dell'adenosintrifosfato (ATP) essenziale per tutti i processi metabolici cellulari; la produzione di ATP permette l'accumulo energetico dell'energia rilasciata dalla catena di trasporto degli elettroni, che genera un gradiente protonico (Mitchell, 1961).

La catena respiratoria (fig. 1.3) è formata da una serie di complessi enzimatici collegati dai trasportatori Coenzima Q e Citocromo *c*. Recentemente è stato dimostrato che queste strutture, organizzate in supercomplessi, non sono rigide ma sembrerebbero modulate dalla componente lipidica (ad esempio la perossidazione dei lipidi sembra disgregarli) .

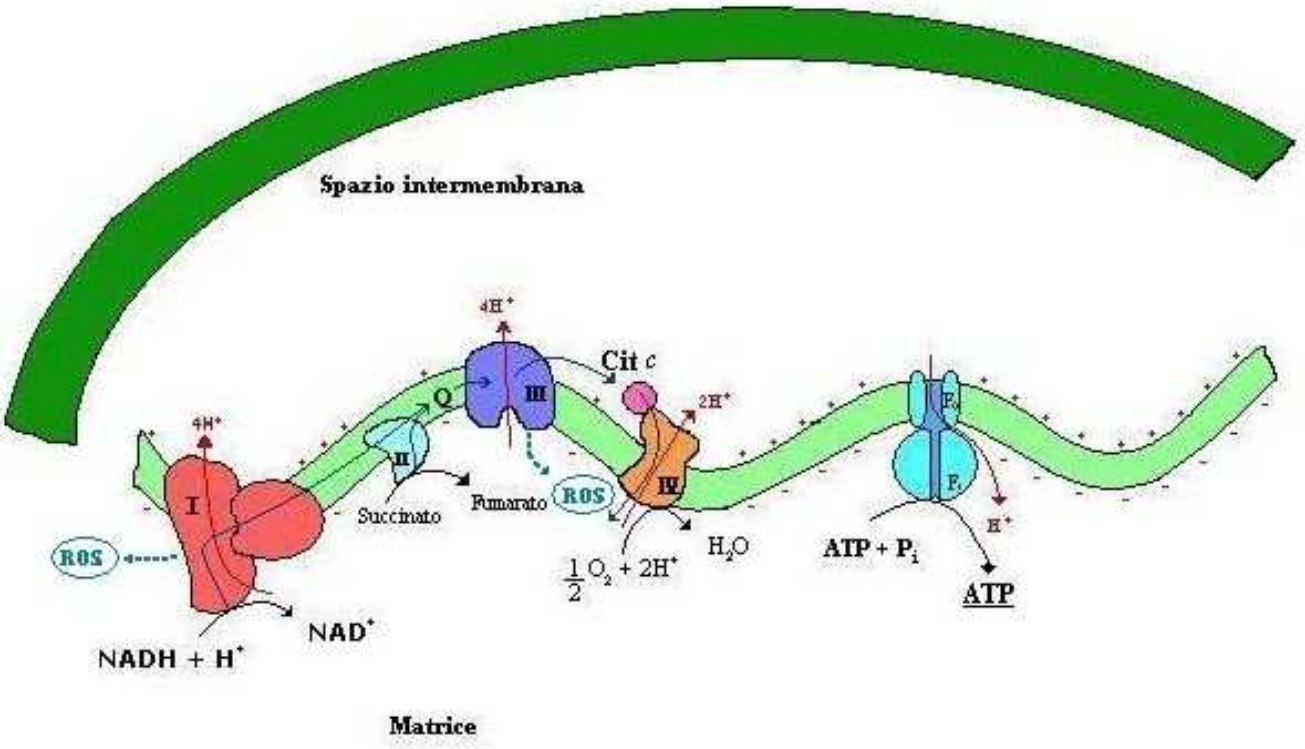


Fig. 1.3 Schema della catena respiratoria che illustra il flusso di elettroni e protoni attraverso i complessi coinvolti. Gli elettroni raggiungono l'ubichinone attraverso i complessi I o II (NADH – CoQ reduttasi e succinato CoQ reduttasi rispettivamente). L'ubichinolo (QH_2) funge da trasportatore mobile e trasferisce gli elettroni al complesso III (ubichinolo citocromo *c* reduttasi) che a sua volta li trasferisce ad un secondo trasportatore mobile, il citocromo *c*. Tramite quest'ultimo gli elettroni giungono al complesso IV (citocromo *c* ossidasi) che catalizza la reazione con l'ossigeno molecolare per formare H_2O. Il flusso di elettroni è accompagnato da una traslocazione di protoni dalla matrice allo spazio intramembrana, che produce un gradiente chimico ΔpH e un gradiente elettrico $\Delta\psi$. Poiché la membrana interna è impermeabile ai protoni, per rientrare questi devono passare attraverso i canali proteici specifici Fo del complesso V (ATPsintasi). Ciò fornisce fornisce l'energia per la sintesi di ATP, catalizzata dal complesso F1.

Sono stati individuati i seguenti complessi fondamentali: il complesso I (NADH-CoQ reduttasi), la cui struttura è ancora poco nota, il complesso III (ubichinolo citocromo *c* reduttasi) e il complesso IV (citocromo *c* ossidasi), di cui invece si è già identificata la struttura. A questi complessi si affianca il complesso II (succinato CoQ reduttasi) che opera in alternativa al complesso I.

L'energia sviluppata dalle reazioni redox che avvengono nei complessi viene accumulata sotto forma di gradiente protonico transmembrana che si ottiene tramite

la traslocazione attiva dei protoni dalla matrice allo spazio intermembrana. Questo gradiente è alla base della sintesi di ATP da parte dell'ATP sintasi (complesso V).

La maggior parte dell' ATP cellulare (90% circa) è prodotto dalla fosforilazione ossidativa (OXPHOS) che ha luogo nella membrana interna dei mitocondri; è proprio nell'ultima fase della fosforilazione che viene utilizzato il gradiente protonico accumulato per produrre ATP a partire da ADP e fosfato inorganico. Il complesso F_1F_0-ATPasi (ATP sintasi) realizza questo passaggio tramite la riossidazione dei coenzimi NADH e $FADH_2$ (possibile grazie alla presenza di sistemi navetta che assicurano il trasporto di equivalenti riducenti del NADH citosolico all'interno dei mitocondri). L'ATP prodotto all'interno dei mitocondri viene trasferito all'esterno tramite il sistema di antiporto della traslocasi dei nucleotidi adenilici (ANT) con contemporaneo ingresso di ADP. Un altro canale, la fosfato traslocasi, assicura che anche il fosfato inorganico P_i raggiunga la matrice mitocondriale. Alterazioni di un solo residuo aminoacidico di questo complesso possono portare conseguenze patologiche molto gravi (sindrome di Leigh). L'ATP sintasi è un complesso enzimatico reversibile, quindi può produrre un gradiente elettrochimico quando la catena respiratoria è malfunzionante. Poiché l'idrolisi dell'ATP può contribuire alla diminuzione di energia quando la disponibilità dell'adenosintrifosfato è già ridotta, questo processo deve essere strettamente controllato. Per questo motivo l'ATP sintasi è sotto il controllo dell'inibitore IF1.

Durante i processi implicati nella produzione del gradiente protonico possono anche formarsi sostanze nocive alla cellula, come l'anione superossido o altre specie

reattive all'ossigeno (ROS), rilasciate all'interno della matrice principalmente dal complesso I e nello spazio intermembrana dal complesso III. La produzione di ROS aumenta quando nel mitocondrio sussiste un eccesso di elettroni, che causa la riduzione permanente dei trasportatori e il passaggio di un elettrone direttamente all'ossigeno, generando così l'anione superossido $O_2^{\cdot-}$. Ciò accade ad esempio se si assumono troppe calorie oppure si riduce l'attività fisica. Per far fronte a questi squilibri il mitocondrio è dotato di sistemi di difesa enzimatici (superossido dismutasi) che convertono immediatamente il superossido in H_2O_2, molecola più stabile del superossido, che può diffondere velocemente nel citosol; H_2O_2 viene quindi convertito in H_2O dalla glutatione perossidasi mitocondriale e citosolica. H_2O_2 in presenza di metalli di transizione può produrre radicali idrossilici altamente reattivi ($^{\cdot}OH$). I centri ferro – zolfo dei complessi proteici mitocondriali sono particolarmente sensibili all'inattivazione causata dai ROS. Pare che la produzione di ROS aumenti drasticamente nei pazienti ammalati di cancro, ma se questa sia la causa della patologia oppure un effetto ancora non è stato chiarito. Sembra comunque chiaro che l'accumulo di ROS giochi un ruolo centrale nelle patologie neurodegenerative correlate all'età (Wallace, 2005).

1.1.3 L'apoptosi

Poiché la maggior parte dei recettori per la regolazione dell'apoptosi (morte cellulare programmata) risiede nella membrana mitocondriale, l'organello ne risulta uno dei principali regolatori (Kroemer, 2003). Nel mitocondrio sono presenti i canali mtPTP (canali a permeabilità transitoria), strutture complesse comprendenti un

canale ionico voltaggio – dipendente (VDAC, situato sulla membrana esterna), un trasportatore di nucleotidi adenilici (ANT, situato sulla membrana interna) e una serie di regolatori dell'apoptosi della famiglia Bcl-2 (*B C*ell *L*eukemia-2), come l'inibitore bcl2 e il promotore bax. In seguito ad appropriati stimoli, i modulatori pro – apoptosi indurranno l'apertura dei canali mtPTP, causando la caduta del potenziale ΔpH e la fuoriuscita delle molecole presenti nello spazio intramembrana, come il citocromo c, l'adattatore ATPasico APAF-1, una serie di pro-caspasi (precursori delle protesi specifiche caspasi), due endonucleasi (endonucleasi G e AIF), alcune molecole inattivatrici degli inibitori citosolici dell'apoptosi (Smac/Diablo) e infine ioni Ca^{2+}, glutatione e radicali liberi. Le prime due molecole formano una struttura chiamata apoptosoma, in grado di reclutare le pro-caspasi e azionare così un meccanismo a cascata che distrugge le proteine citoplasmatiche. Le endonucleasi sono trasportate nel nucleo, dove si attivano e degradano il DNA nucleare. Tutte le altre molecole contribuiscono al processo apoptotico, consentendo la degradazione totale delle strutture interne alla cellula, ma non della membrana cellulare. Oltre che dagli stimoli esterni, l'apertura dei canali può essere causata da stress ossidativi, diminuzione di ADP o del potenziale transmembrana. Recentemente è stata ipotizzata un'associazione tra mutazioni del DNA mitocondriale e fenomeni di apoptosi legati all'invecchiamento (Chomyn et al, 2003).

1.2 Il DNA mitocondriale

Ciascuna cellula umana (salvo poche eccezioni) contiene da una a diverse centinaia di mitocondri, ciascuno caratterizzato, a seconda del tipo cellulare, da un numero variabile di copie (>1000) di DNA mitocondriale (mtDNA). L'alto numero di copie risultanti fa sì che, nonostante le piccole dimensioni, il genoma mitocondriale possa arrivare a costituire fino allo 0.3% dell'intero DNA cellulare.

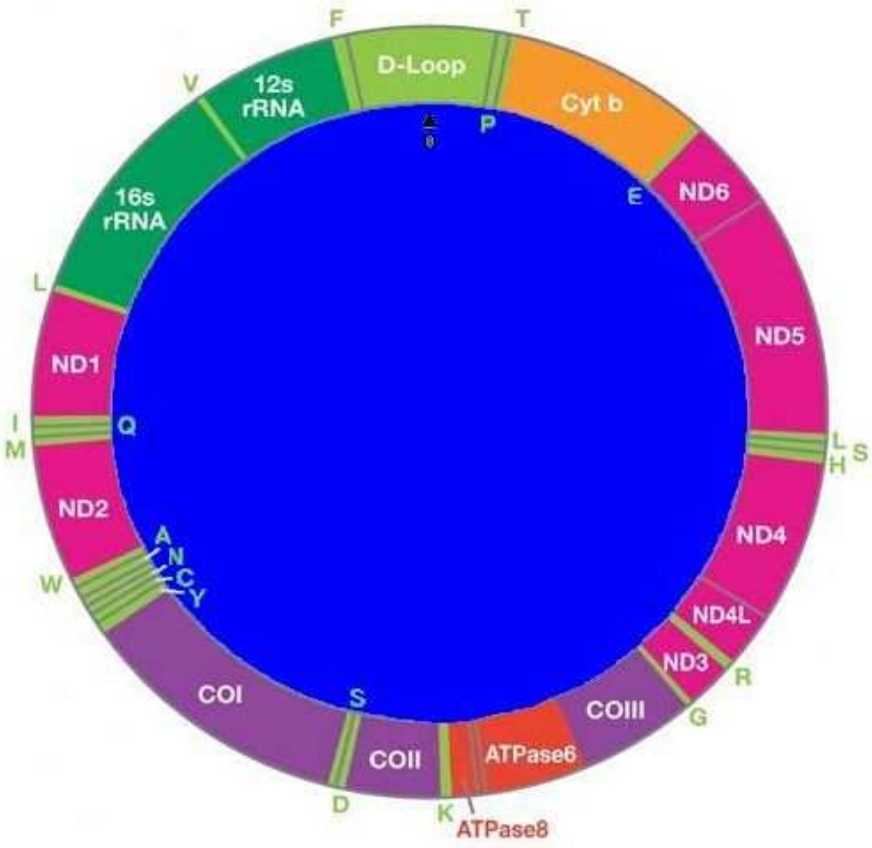


Fig. 1.4 Mappa del DNA mitocondriale. Sono riportati i 37 geni codificati dai due filamenti, come riportato nella Tabella I.1; è indicato inoltre il D-loop.

Il DNA mitocondriale umano è una molecola circolare a doppia elica costituita da 16569 coppie di basi che si trova all'interno della matrice del mitocondrio (fig.1.4). I due filamenti che la compongono sono chiamati H (filamento pesante ricco di residui guaninici) e L (filamento leggero ricco di residui citosinici). Le dimensioni ridotte e l'alto numero di copie hanno consentito di determinarne la sequenza già nel 1981 (Anderson et al, 1981) e di studiarne approfonditamente l'organizzazione. In seguito vari studi di sequenziamento hanno permesso di correggere gli errori originari e di ottenere una sequenza più aggiornata (Andrews et al, 1999). Il DNA mitocondriale è quasi tutto codificante, i suoi geni a differenza di quelli nucleari sono privi di introni, le sequenze intergeniche sono poche e brevi e talvolta le regioni codificanti si sovrappongono alle estremità.

Le uniche eccezioni a questa situazione di estrema compattezza sono costituite da:

- una sequenza lunga 1112 bp, localizzata tra i geni che codificano per il $tRNA^{pro}$ e il $tRNA^{phe}$, denominata regione di controllo o regione del D-loop (displacement loop), all'interno della quale si trovano l'origine di replicazione del filamento pesante, i promotori della trascrizione di entrambi i filamenti e la sequenza di terminazione TAS; all'interno del D-loop si possono identificare due regioni ipervariabili (HV1 e HV2) di circa 400bp ciascuna, che contengono ben il 92% dei siti polimorfici di tutta la regione, e una terza regione ancora poco conosciuta denominata HV3.
- la regione V, lunga 121 bp, localizzata tra il gene codificante per la citocromo *c* ossidasi II e il $tRNA^{lys}$.

Filamento	Gene	Posizione
H	tRNA Fenilalanina (**F**)	577 – 647
H	rRNA **12S**	648 – 1601
H	tRNA Valina (**V**)	1602 – 1670
H	rRNA **16S**	1671 – 3228
H	tRNA Leucina 1(**L** - UUA/G)	3230 – 3304
H	NADH deidrogenasi 1 (**ND1**)	3307 – 4262
L	tRNA Glutammina (**Q**)	4329 – 4400
H	tRNA Metionina (**M**)	4402 – 4469
H	NADH deidrogenasi 2 (**ND2**)	4470 – 5511
H	tRNA Triptofano (**W**)	5512 – 5579
L	tRNA Alanina (**A**)	5587 – 5655
L	tRNA Asparagina (**N**)	5657 – 5719
L	tRNA Cisteina (**C**)	5761 – 5826
L	tRNA Tirosina (**Y**)	5826 – 5891
H	Citocromo *c* ossidasi I (**COI**)	5904 – 7445
L	tRNA serina 1 (**S** – UCN)	7446 – 7516
H	tRNA Acido Aspartico (**D**)	7518 – 7585
H	Citocromo *c* ossidasi II (**COII**)	7586 – 8269
H	tRNA Lisina (**K**)	8295 – 8364
H	ATP sintasi Fo subunità 8 (**ATPasi8**)	8366 – 8572
H	ATP sintasi Fo subunità 6 (**ATPasi6**)	8527 – 9207
H	Citocromo *c* ossidasi III (**COIII**)	9207 – 9990
H	tRNA Glicina (**G**)	9991 – 10058
H	NADH deidrogenasi 3 (**ND3**)	10059 – 10404
H	tRNA Arginina (**R**)	10405 – 10469
H	NADH deidrogenasi 4L (**ND4L**)	10470 – 10766
H	NADH deidrogenasi 4 (**ND4**)	10760 – 12137
H	tRNA Istidina (**H**)	12138 – 12206
H	tRNA Serina 2 (**S** – AGY)	12207 – 12265
H	tRNA Leucina 2 (**L** – CUN)	12266 – 12336
H	NADH deidrogenasi 5 (**ND5**)	12337 – 14148
L	NADH deidrogenasi 6 (**ND6**)	14149 – 14673
L	tRNA Acido Glutammico (**E**)	14674 – 14742
H	Citocromo b (**Cytb**)	14747 – 15887
H	tRNA Treonina (**T**)	15888 – 15953
L	tRNA Prolina (**P**)	15955 – 16023

Tab. I.1 Elenco dei geni mitocondriali. Per ogni gene è indicato il nome esteso e la sigla corrispondente alla mappa in Figura 1.4, il filamento codificante, il nucleotide iniziale e quello finale.

Il DNA mitocondriale codifica per 13 componenti proteiche, tutte coinvolte nel processo di produzione di ATP attraverso la fosforilazione ossidativa (OXPHOS), 2

RNA ribosomali (rRNA 12S e 16S) e 22 RNA di trasporto (tRNA) necessari e sufficienti per la traduzione del messaggio informazionale del mtDNA (tab. I.1).

La fosforilazione ossidativa ha luogo sulla membrana interna dei mitocondri mediante 5 complessi enzimatici multiproteici (complessi I, II, III, IV e V) i cui geni sono ripartiti tra il DNA mitocondriale e il DNA nucleare. Le proteine codificate da quest'ultimo vengono riconosciute grazie alla presenza di una sequenza leader all'estremità N-terminale e trasportate nei mitocondri.

Il DNA mitocondriale codifica per 7 (ND1, ND2, ND3, ND4, ND4L, ND5, ND6) dei 43 polipeptidi del complesso I (NADH-Q ossido-reduttasi o NADH deidrogenasi); un polipeptide (citocromo b), degli 11 che costituiscono il complesso III (citocromo *c* ossido-reduttasi); 3 polipeptidi (COI, COII, COIII) dei 13 del complesso IV (citocromo *c* ossidasi) ed infine 2 polipeptidi (ATPasi 6 ed ATPasi 8) dei 16 che formano il complesso V (ATP sintasi). I due rRNA, 13 tRNA, le subunità ND1, ND2, ND3, ND4, ND4L, ND5 della NADH deidrogenasi, il citocromo b, le subunità I, II e III della citocromo *c* ossidasi e le subunità 6 e 8 dell'ATPsintasi sono codificate dal filamento pesante (più ricco in purine), mentre il filamento leggero (più ricco in pirimidine) codifica solo per 8 tRNA e per la subunità 6 della NADH deidrogenasi. C'è da sottolineare poi che i tRNA presentano una caratteristica disposizione punteggiata.

Il DNA nucleare codifica per tutte le altre proteine mitocondriali, incluse le subunità della polimerasi γ (POLG), le componenti dell'RNA polimerasi e i fattori di trascrizione mtTFA. Il numero ridotto di polipeptidi, codificati dal genoma

mitocondriale, rappresenta, infatti, una minima frazione di tutte le proteine necessarie per lo svolgimento delle funzioni mitocondriali. La maggior parte di queste è codificata dal DNA nucleare e viene esportata ai mitocondri in un secondo tempo.

Il DNA mitocondriale costituisce un'eccezione all'universalità del codice genetico. Nei mammiferi molti cambiamenti riguardano codoni coinvolti nell'inizio o nella terminazione della sintesi proteica. I codoni AGA e AGG sono entrambi letti come segnali di terminazione invece che come codoni per l'arginina; il codone AUA è letto come codone per la metionina e non per l'isoleucina. Le differenze tra i codici di lettura del DNA nucleare e mitocondriale sono riassunti nella tabella I.2. L'unica modifica comune a tutti i DNA mitocondriali, è che il codone UGA ha lo stesso significato di UGG e quindi codifica per il triptofano invece che essere un codone di stop.

Codone	**Codice mtDNA mammiferi**	**Codice mtDNA lievito**	**Codice Universale**
UGA	Trp	Trp	STOP
AUA	Met	Met	Ile
CUA	Leu	Thr	Leu
AGA	STOP	Arg	Arg
AGG	STOP	Arg	Arg

Tab. I.2 Variazioni del codice genetico nel DNA mitocondriale.

I mitocondri possono avere un codice genetico diverso dal resto della cellula perché il DNA mitocondriale codifica per una serie distinta di tRNA. I 22 tRNA codificati dal mtDNA sono sufficienti per leggere i 13 geni codificanti le proteine mitocondriali. Ciascuno dei 22 tRNA è infatti in grado di leggere una famiglia di due o quattro codoni.

La replicazione delle molecole di mtDNA avviene a partire da due diverse origini, che si trovano rispettivamente sul filamento pesante (O_H) e sul filamento leggero (O_L). Secondo il modello tradizionale di Clayton la replicazione inizia con la sintesi della catena pesante a partire da vari punti precisi all'interno del D-loop. L'allungamento della catena nascente porta al distacco dei due filamenti, finché non viene scoperta l'origine di replicazione del filamento leggero; a questo punto può aver inizio anche la replicazione di quest'ultimo che avviene in direzione opposta (Clayton, 1982). L'inizio della replicazione avviene a livello dell'origine O_H utilizzando un innesco ad RNA, generato dalla trascrizione del filamento leggero. Quest'ultimo infatti viene tagliato da una RNAsi codificata nel nucleo in alcune posizioni specifiche ben conservate ed il risultante 3′-OH viene utilizzato dalla DNA polimerasi γ come *primer* per sintetizzare un DNA 7S che termina alla sequenza associata alla terminazione (TAS), situata all'estremità della regione di controllo. La sequenza TAS si lega ad un fattore specifico in grado di regolare le pause nella replicazione. Questa regione neosintetizzata espone il filamento pesante parentale in modo da creare l'ansa D (displacement loop). All'interno del D-loop esistono numerosi siti di origine per il filamento H, ma solo alcune di queste catene nascenti

vengono effettivamente estese per dare una replicazione completa. La maggior parte va invece a costituire il DNA7S. Recentemente Fish e coll. Hanno evidenziato l'esistenza di una nuova origine nella posizione 57, che sembrerebbe essere la "vera" origine di replicazione, perché nessuno dei filamenti nascenti da questa posizione termina prematuramente (fig. 1.5).

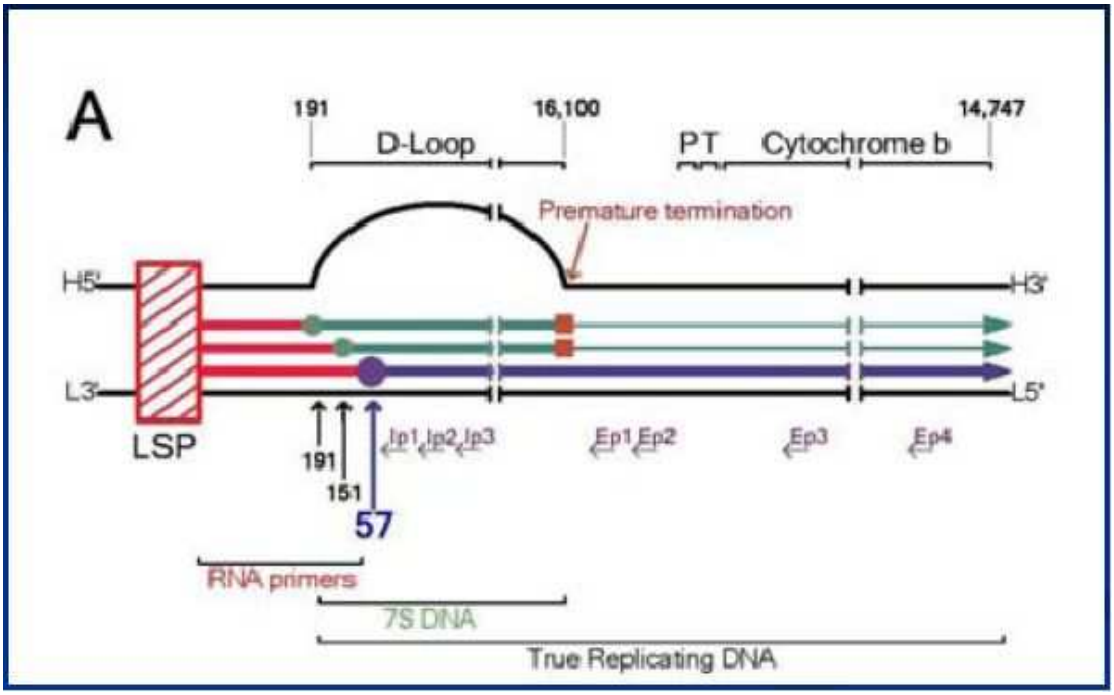


Fig. 1.5 Schema rappresentante l'inizio della replicazione del filamento H del DNA mitocondriale. I numeri indicano la posizione dei nucleotidi, Ip1, Ip2, Ip3, Ep1, Ep2, Ep3 Ep4 rappresentano i *primer*s utilizzati nel lavoro di Fish et al. (2004) per analizzare la zona del D-loop che coinvolge l'inizio della replicazione; è evidenziata la "vera" origine di replicazione (nucleotide 57).

Il lavoro di Fish sembra anche dimostrare che in condizioni di stress, in cui la cellula deve aumentare i livelli di replicazione, anche il DNA7S può proseguire l'allungamento per fornire nuovi DNA integri (Fish et al, 2004) .

La trascrizione del mtDNA produce RNA policistronici che successivamente vengono processati per dare origine a tRNA e mRNA maturi. Per iniziare la

trascrizione sui due promotori localizzati nel D-loop (P_H e P_L), l'RNA polimerasi specifica (polRMT) richiede il fattore di trascrizione TFAM (una proteina con due domini di legame al DNA ed una coda all'estremità C-terminale, essenziale per la trascrizione) ed un altro dei due fattori B1 o B2 (tutti e tre codificati del nucleo). Recenti ricerche hanno evidenziato che TFAM induce dei cambiamenti strutturali nel promotore del filamento leggero; tali cambiamenti sembrano indispensabili per il corretto riconoscimento del promotore da parte dalla polRMT (Fernandez-Silva et al, 2006).

1.3 L'ereditarietà e la variabilità

1.3.1 Ereditarietà materna

Studi effettuati sul mtDNA utilizzando gli RFLPs (*Restriction Fragment Lenght Polymorphisms*) hanno mostrato come il DNA mitocondriale venga ereditato per via materna e come invece quello paterno non si osservi nella progenie. Tra le varie ipotesi che sono state formulate, alcune mettono in dubbio la vitalità dei mitocondri degli spermatozoi, che anche riuscendo ad entrare nell'oocita, dovrebbero essere fortemente danneggiati nelle loro funzioni a causa dell'eccessiva produzione di radicali liberi per la produzione di energia durante il movimento dello spermatozoo stesso. E' stata anche ipotizzata la presenza dell'ubiquitina all'interno degli oociti, una proteina che degrada il materiale esterno alla cellula e quindi potrebbe attaccare anche il DNA mitocondriale paterno (Sutovsky et al, 2000). Una delle ipotesi più

accreditate si basa su un calcolo puramente probabilistico; infatti, nella maggior parte dei mammiferi, incluso l'uomo, al momento della fecondazione gli spermatozoi completi della coda, contenente i mitocondri, penetrano all'interno dell'oocita. Ogni spermatozoo contiene al momento della penetrazione da 50 a 60 mitocondri: un numero molto piccolo se si considera che nell'oocita maturo sono presenti da 100.000 a 200.000 copie di mtDNA.

L'ereditarietà uniparentale è stata messa in discussione recentemente, così come l'assenza di ricombinazione.

Per quanto riguarda il primo quesito, in realtà finora, mentre nei topi, nella Drosophila e nei mitili sono state trovate prove della presenza di poche molecole di mtDNA paterno (10^{-4} per generazione) nella progenie, nell'uomo non sono mai state trovate prove dirette della presenza di materiale paterno, se non in un singolo caso, dove però si è analizzato un paziente con miopatia mitocondriale e con mtDNA fortemente mutato che, a livello del tessuto muscolare, sembrerebbe più vicino a quello paterno (Schwartz et al, 2002). I dati che provano definitivamente che il DNA mitocondriale è ereditato solo per via materna, e che il caso di Schwartz è solo un'eccezione, sono stati ottenuti recentemente da Nishimura, il quale ha eseguito una serie di osservazioni su alcuni spermatozoi viventi, dimostrando che sussiste una graduale diminuzione dei mitocondri durante la spermatogenesi e una rapida digestione dei pochi mtDNA rimanenti subito dopo la fecondazione, digestione che avviene ancor prima che le strutture mitocondriali siano completamente distrutte,

forse per impedire la trasmissione di mtDNA spermatici dannosi alla progenie (Nishimura et al, 2006).

Per quanto riguarda la ricombinazione, non sono mai stati osservati fenomeni ricombinativi nel mtDNA umano (Innan e Nordborg, 2002; Bandelt et al, 2005): ciascuna molecola si trasmette quindi invariata dalla madre alla progenie, salvo mutazioni casuali. Se il DNA mitocondriale ne fosse soggetto, tutte le ipotesi riferite alla nostra storia evolutiva finora effettuate con il modello dell'ereditarietà materna sarebbero prive di fondamento. In tale modello non sono previste né la segregazione né la ricombinazione del mtDNA, pertanto l'ascendenza di un individuo anche dopo molte generazioni può essere ricondotta ad un solo antenato, analogamente a quanto avviene con il cromosoma Y (utilizzato infatti per seguire tramite la linea paterna la storia evolutiva dell'uomo; Ke et al, 2001) ed a differenza di quanto succede, invece, per il DNA nucleare.

1.3.2 Tasso evolutivo e mutazioni

Nonostante il DNA mitocondriale sia altamente conservato nel regno animale per struttura, dimensioni ed organizzazione, è stata osservata una notevole variabilità a livello della sequenza nucleotidica. Si è stimato che il tasso evolutivo del mtDNA è di circa un ordine di grandezza superiore a quello dei geni nucleari a singola copia (in media circa 10-20 volte più veloce; Wallace et al, 1987). Ciò è diretta conseguenza del tasso di mutazione, secondo la formula: $\mathbf{E = M \times F}$, dove F rappresenta la frazione fissata di mutazioni ed M il tasso a cui avvengono le mutazioni. Il parametro M risulta maggiore nel DNA mitocondriale a causa dello sbilanciamento nella

concentrazione dei quattro dNTPs (Kaguni, 2004) nonché della maggiore quantità dei cicli di replicazione, durante i quali è possibile che si generino un maggior numero di errori. L'elevata frequenza di mutazione di un gene mitocondriale rispetto ad un gene nucleare, risulta di notevole utilità negli studi evolutivi relativi all'origine della nostra specie e delle popolazioni umane ed è causa di numerose patologie mitocondriali più o meno gravi (vedi paragrafo 1.3.3). Oltre ad una percentuale di mutazioni più alta, nel DNA mitocondriale si rileva anche un aumento della frazione F; infatti poiché la dimensione effettiva della popolazione, basandosi sullo studio del solo DNA mitocondriale è un quarto di quella reale (una coppia di riproduttori diploidi contiene quattro genomi aploidi nucleari ma solo un tipo di mtDNA), l'effetto della deriva genetica (inversamente proporzionale alla dimensione della popolazione considerata) risulta quadruplicato rispetto alle osservazioni riguardanti il DNA nucleare.

In accordo con la teoria endosimbiontica di Lynn-Margulis, i mitocondri possono essere considerati come una popolazione intracellulare di microrganismi. Quando insorge una mutazione in uno degli mtDNA presenti all'interno delle cellule, si viene a creare una situazione in cui coesistono molecole mutate e normali, situazione nota come **eteroplasmia.** I DNA mitocondriali di una cellula eteroplasmica che si divide, vanno incontro ad un processo noto come segregazione replicativa in cui la segregazione di molecole di tipo selvatico o di tipo mutato nelle cellule figlie, dipende esclusivamente da fenomeni casuali di deriva genetica. Il susseguirsi di tali eventi può portare nel tempo alla fissazione differenziale dei due tipi di mtDNA in cellule diverse dello stesso tessuto oppure in diversi tipi cellulari.

Infine se c'è eteroplasmia anche nelle cellule della linea germinale materna, si potrà riscontrare la fissazione differenziale dei diversi mtDNA in diversi discendenti della stessa madre eteroplasmica.

L'esistenza di una condizione eteroplasmica per singole sostituzioni nucleotidiche, piccole inserzioni o delezioni è stata osservata nei vertebrati superiori, ma sembra essere più frequente in alcuni vertebrati inferiori ed in *Drosophila mauritiana*. Nell'uomo, in particolare nella regione di controllo, è stato osservato uno stato di eteroplasmia per alcune mutazioni neutrali e l'analisi molecolare dei discendenti di donne eteroplasmiche ha spesso rivelato omoplasmia nelle generazioni successive. La condizione di eteroplasmia è relativamente comune nell'uomo in soggetti che presentano un quadro clinico di una certa gravità, dovuto ad esempio alla presenza di mutazioni altamente dannose nei geni dei tRNA; gli effetti di tali mutazioni sono così devastanti che possono essere il risultato solo di eventi mutazionali molto recenti; quando il livello dell'eteroplasmia è elevato per mutazioni dannose, infatti, queste ultime vengono perse in poche generazioni. Le patologie legate a mutazioni eteroplasmiche hanno gravità diversa a seconda della proporzione di mtDNA normali e mutati nei tessuti affetti di un individuo.

Non è ancora del tutto chiaro il meccanismo mediante il quale viene mantenuta l'omoplasmia, mentre risulta evidente un'estrema rapidità nel fissare o perdere le nuove mutazioni. Una possibile spiegazione tiene conto del fatto che le circa 100.000 molecole di mtDNA presenti in un oocita maturo di mammifero derivano da un numero molto ridotto di molecole capostipiti, forse addirittura solo cinque copie.

Questa espansione clonale della popolazione di mtDNA nell'oocita durante l'oogenesi crea una situazione particolarmente idonea alla deriva genetica ed a fenomeni analoghi a quello definito "effetto del fondatore".

Le mutazioni del mtDNA possono essere di tre tipi: mutazione per sostituzione di una coppia di basi (mutazioni puntiformi), per inserzione, per delezione.

Le sostituzioni nucleotidiche si possono suddividere in transizioni e transversioni. Al primo gruppo appartengono le sostituzioni adenina - guanina (A↔G) e citosina - timina (C↔T) ovvero purina – purina e pirimidina – pirimidina rispettivamente. Alle transversioni appartengono le sostituzioni di una purina con una pirimidina e viceversa (ad esempio A↔C o G↔T). Complessivamente esistono due tipi di transizione e quattro tipi di transversione. Gli errori nella replicazione, che sono alla base delle mutazioni puntiformi sono dovuti principalmente alla proprietà chimica delle basi azotate di assumere due diverse forme con diversa stabilità (tautomerismo cheto – enolico o amino – iminico). Quando le basi si trovano nello stato più stabile, cheto – amminico, si appaiano secondo le regole di Watson – Crick; al contrario, il tautomerismo può generare sostituzioni puntiformi nella classe delle transizioni, consentendo appaiamenti non canonici, come ad esempio G^{enolo} – T o C^{imino} – A. Tra le cause che danno origine a transversioni troviamo: a) la rotazione del legame glicosidico di una purina in un orientamento energeticamente sfavorito; b) reazioni di ossidazione ad opera di radicali dell'ossigeno che trasformano la guanina in 8-idrossiguanina, causandone l'appaiamento preferenziale con l'adenina piuttosto

che con la citosina. La stima delle probabilità che questi fenomeni avvengano è di $10^{-4} – 10^{-5}$ per le transizioni e di $10^{-5} – 10^{-6}$ per le transversioni, ma nella realtà, sebbene il numero delle transizioni sia effettivamente superiore a quello delle transversioni, accurati meccanismi di individuazione e riparo degli errori mantengono un tasso di mutazione di $10^{-8} – 10^{-10}$ nel DNA nucleare, tasso che risulta invece più elevato nel DNA mitocondriale (vedi oltre). Questo, in effetti, possiede minori meccanismi di riparo rispetto al primo e la polimerasi mitocondriale, dal canto suo, è priva della capacità di correggere le bozze (attività "proof reading") durante la replicazione.

Le conseguenze di ciascuna mutazione sono determinate dalla sua natura molecolare e del suo verificarsi nelle cellule somatiche o nella linea germinale (Wallace et al, 1994; Ruiz-Pesini et al, 2004; Lee et al, 2009), inoltre possono variare anche in funzione dello stadio del ciclo vitale dell'individuo portatore (Jakupciak et al, 2008). Le mutazioni del mtDNA nella linea germinale sono uno strumento particolarmente idoneo per lo studio dell'evoluzione umana e possono essere sia neutrali che dannose. Le mutazioni neutrali si sono accumulate sequenzialmente e indipendentemente sulle diverse linee di radiazione materne e permettono di distinguere individui o popolazioni umane che abbiano tempi di separazione ridotti da un punto di vista evolutivo. Le mutazioni dannose o potenzialmente letali vengono invece eliminate semplicemente per effetto della selezione naturale tranne in alcuni casi in cui vengono mantenute in stato di eteroplasmia, ma sono comunque il prodotto di eventi mutazionali recenti. Le mutazioni somatiche pur non essendo utili

ai fini dello studio dell'evoluzione umana, possono giocare un ruolo notevole nel modulare l'espressione di patologie gravi e nei normali processi di invecchiamento. Queste mutazioni possono essere dovute sia a riorganizzazioni genomiche, sia a singole sostituzioni nucleotidiche, entrambe comunque alterano la normale capacità di attuare la fosforilazione ossidativa in particolar modo nei tessuti post-mitotici. L'effetto di una minore capacità di produrre energia da parte della catena respiratoria si manifesta con la comparsa precoce dei sintomi clinici soprattutto in quei tessuti che devono soddisfare un maggior fabbisogno energetico come il cervello, il sistema nervoso, i muscoli cardiaci, la muscolatura scheletrica, i reni, il sistema endocrino ed il fegato.

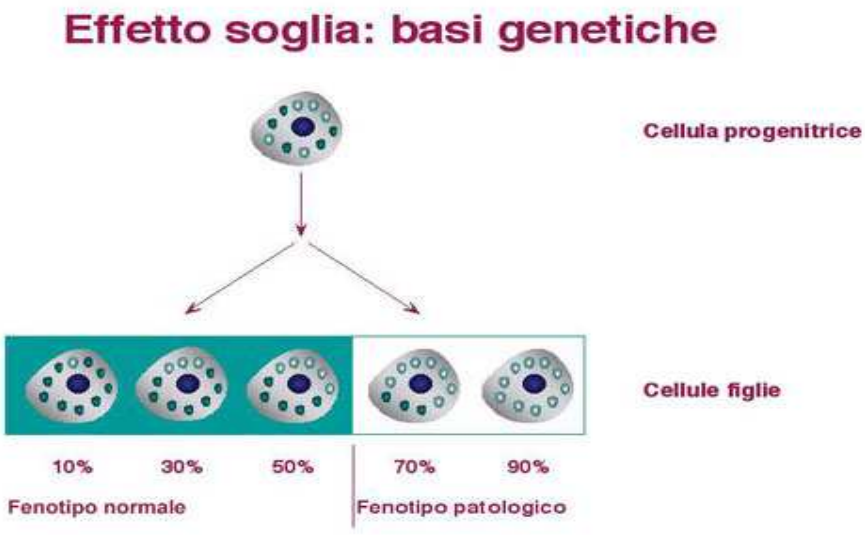


Fig. 1.6 Effetto soglia. La figura illustra l'effetto casuale con cui i mitocondri portatori della mutazione (in bianco) possono segregare nelle cellule figlie, determinando se il soggetto sarà sano o svilupperà la patologia con intensità variabile.

Un individuo sano che possiede un genotipo normale per quanto riguarda la capacità di attuare la fosforilazione ossidativa, dovrà accumulare una quantità molto grande di mutazioni somatiche prima di raggiungere la soglia energetica minima richiesta da un qualsiasi tessuto (circa il 60 %) e dunque manifestare i sintomi clinici (fig. 1.6). In questo modo, tali individui non risentiranno di disfunzioni energetiche e non mostreranno disfunzioni cliniche se non in età avanzata; sembra infatti che l'accumulo di mutazioni somatiche sia una normale componente del processo di invecchiamento, dovuto probabilmente all'accumulo di radicali dell'ossigeno all'interno dei mitocondri (vedi sopra). Al contrario, un individuo che ha ereditato dalla madre un genotipo mitocondriale che presenta numerose mutazioni dannose, avrà una ridotta capacità bioenergetica iniziale e dunque supererà prima la soglia minima di richiesta energetica per ogni tessuto, manifestando un quadro clinico più o meno grave a seconda delle mutazioni coinvolte.

1.3.3 Mitocondriopatie derivanti da mutazioni del DNA mitocondriale umano

Dato il ruolo vitale dei mitocondri per la produzione cellulare di energia, molte malattie genetiche sono associate con mutazioni dell'mtDNA (revisione di Wallace, 1999). Poiché i mitocondri sono presenti in tutti i tessuti, le malattie mitocondriali possono colpire qualsiasi organo. Più spesso, però, interessano il muscolo ed il cervello data la maggiore richiesta di energia di questi tessuti, specie durante lo sviluppo. Per questo motivo, le malattie mitocondriali sono spesso definite come

encefalo-mio-patie mitocondriali. La figura 1.7 illustra alcune delle principali patologie mitocondriali, distinguendo i geni coinvolti (Di Mauro, 2004).

Le variazioni fenotipiche, a livello di proteine, ribosomi o tRNA codificati dal genoma mitocondriale, possono potenzialmente alterare la produzione cellulare dell'energia. Si ritiene che tali mutazioni puntiformi possano portare ad un fenotipo difettoso ("leaky"), nel quale la proteina od il tRNA alterati comunque funzionano, sia pure in maniera parziale (Scheffler, 1999). Tuttavia la riduzione di efficienza della fosforilazione ossidativa ("OXPHOS") può avere conseguenze mediche signficative che si possono manifestare in maniera differenziale nel tempo (alcune malattie da mutazione puntiforme sono, infatti, ad insorgenza tardiva) o tra i tessuti (Scheffler, 1999). Variazioni nell'assetto metabolico possono a loro volta indurre l'apoptosi delle cellule, dal momento che la permeabilità mitocondriale gioca un ruolo centrale nella via della morte cellulare apoptotica (revisione di Li uet al, 1997).

Le malattie mitocondriali sono state descritte fin dai primi anni sessanta, quando studi istochimici e strutturali rivelarono un'eccessiva proliferazione di mitocondri (spesso anormali) nel tessuto muscolare di pazienti con debolezza e intolleranza all'esercizio fisico. Poiché le aree interessate apparivano rosse in seguito alla colorazione Gomari (fig. 1.8), le fibre furono chiamate RRF (*R*agged-*R*ed *F*ibers) e considerate indicative delle patologie mitocondriali. In realtà, molte, ma non tutte le mitocondriopatie presentano questa caratteristica.

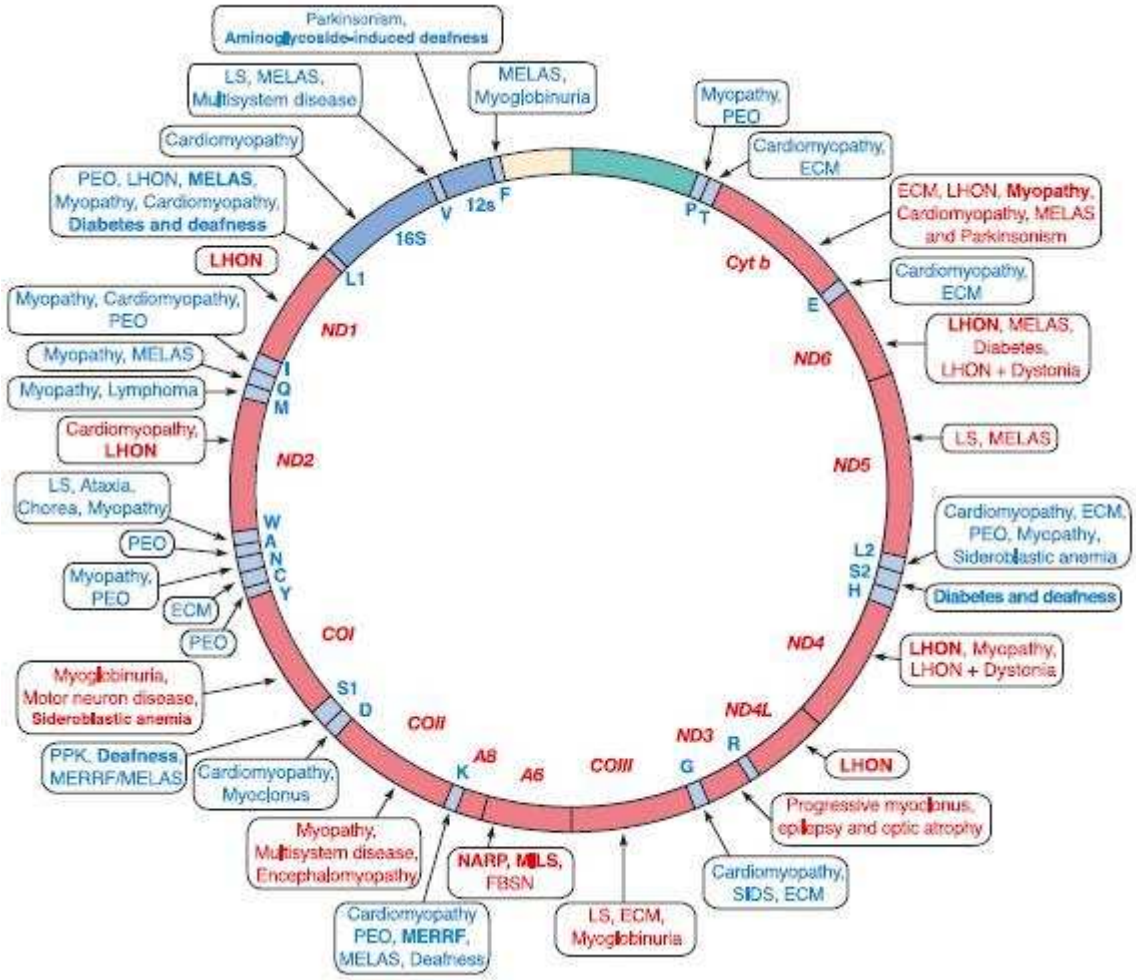


Fig. 1.7 Mappa morbida del DNA mitocondriale umano. La mappa illustra i geni presenti sul DNA e le malattie più comuni associate a difetti nella sintesi di proteine (in blu) e a difetti nei geni codificanti proteine (in rosso) (Di Mauro, 2004).

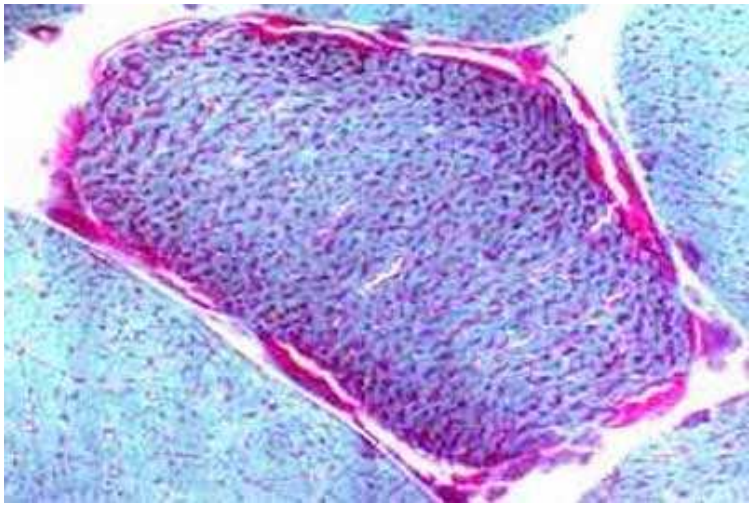


Fig. 1.8 Fibre rosse sfilacciate. La figura mostra la tipica colorazione assunta dagli accumuli periferici di mitocondri in cellule muscolari di pazienti con mitocondriopatie.

Esiste oltre un centinaio di malattie genetiche mitocondriali, definite, o associate o implicate nelle mutazioni dell'mtDNA (Kogelnik et al, 1998). Il sito web di MitoMap contiene una lista, che viene continuamente aggiornata ed implementata, di mutazioni dell'mtDNA malattia-associate/implicate (http://www.mitomap.org).

Le mutazioni del mtDNA possono essere divise in due categorie; della prima fanno parte quelle che coinvolgono la sintesi delle proteine (tRNA e rRNA), della seconda quelle che coinvolgono una delle 13 subunità della catena respiratoria codificate dal DNA mitocondriale stesso. Alcune di queste patologie sono illustrate nella tabella I.3.

Gene	Difetto biochimico	Fenotipo clinico
tRNALeu(UUR)	↓ sintesi proteica	MELAS
tRNALys	↓ sintesi proteica	MERRF
Altri tRNA	↓ sintesi proteica	Fenotipi multipli
ATPasi6	↓ sintesi di ATP	NARP/MILS
ND1,ND4, ND6	↓ funzionalità complesso I	LHON
ND1,ND4	↓ funzionalità complesso I	miopatia
Cyt b	↓ funzionalità complesso III	miopatia
COX III	↓ funzionalità complesso IV	miopatia

Tab. I.3 Effetti delle mutazioni nel DNA mitocondriale: la Tabella mostra alcune tra le più comuni patologie derivate da mutazioni puntiformi del DNA mitocondriale; le patologie associate a difetti di Cytb e COXIII causano malattie multisistemiche.

1.3.3.1 Difetti nella sintesi proteica

Singole delezioni del mtDNA sono state associate a tre condizioni tipiche: sindrome di Pearson (malattia infantile caratterizzata da anemia e disfunzioni

pancreatiche, che porta rapidamente alla morte), sindrome di Kearns–Sayre (KSS, malattia multisistemica che insorge prima dei vent'anni e porta alla progressiva atrofia ottica, alla retinopatia, al blocco cardiaco ed è frequentemente associata ad atassia, demenza e diabete mellito), oftalmoplegia progressiva ereditaria (PEO). Oltre alle singole delezioni in alcuni pazienti sono state osservate delezioni di 5 kb, mentre duplicazioni sono comuni tra i pazienti con diabete mellito o KSS. Entrambi i riarrangiamenti sono rari e solitamente sono ereditati per via materna.

Più comuni sono le mutazioni puntiformi, osservate in patologie come la MELAS (encefalopatia mitocondriale, acidosi lattica e colpi apoplettici) che si presenta in età giovanile. Tra le mutazioni ad essa associate c'è la transizione A3243G, a carico del tRNA della leucina. Un'altra patologia dovuta a mutazioni puntiformi è la MERRF (epilessia mioclonica con fibre rosse sfilacciate); anche in questo caso è coinvolto un tRNA, quello della lisina (A8344G, T8356C, G8363A). Recentemente Mancuso ha individuato una paziente con una mutazione nel $tRNA^{Phe}$ (G611A) (Mancuso et al, 2004). Poiché difetti nei tRNA coinvolgono vari sistemi (ottico, cardiaco, nervoso, gastrointestinale) quando un individuo presenta sintomi che ne coinvolgono più d'uno è lecito sospettare una patologia mitocondriale.

1.3.3.2 Difetti nei geni codificanti proteine

Sono due le patologie principali che rientrano in questa categoria.

La prima è una combinazione di due malattie, la NARP (neuropatia, atassia, retinite pigmentosa) e la MILS (Sindrome di Leigh ereditaria, un'encefalopatia infantile). Entrambe le patologie sono dovute ad una mutazione nel gene codificante la subunità enzimatica ATP6 nella posizione 8993. Sono state riscontrate una transizione T→C e una transversione T→G (più grave). La mutazione T8993G è un eccellente esempio del fenomeno dell'effetto soglia e dell'importanza del livello di eteroplasmia del fenotipo mutante; infatti, quando il paziente presenta un'eteroplasmia del 70% circa, l'espressione clinica è la NARP, mentre se l'eteroplasmia è più consistente (circa 90%), insorgono complicazioni come la sindrome del "bambino flaccido" e la sindrome di Leigh.

La seconda patologia è la LHON (neuropatia ottica di Leber), caratterizzata da perdita della vista in età giovanile (specialmente nei maschi) dovuta ad atrofia ottica bilaterale. Sono state individuate tre mutazioni primarie ad essa associate, tutte nei geni ND del complesso I (G11778A in ND4, G3460A in ND1 e T14484C in ND6).

1.3.4 Mitocondriopatie derivanti da mutazioni del DNA nucleare

Come già indicato precedentemente, oltre il 90% delle proteine mitocondriali sono espressione di geni nucleari. Tuttavia, le patologie causate da mutazioni in questi geni, coinvolti nel funzionamento ed assemblaggio delle subunità enzimatiche della catena respiratoria, sono poco conosciute. Le recenti ricerche riguardanti le malattie neurodegenerative ereditarie confermano che un gran numero di queste

ultime derivano dalla mutazione di proteine codificate nel nucleo. Una serie di funzioni all'interno del mitocondrio deriva direttamente o indirettamente da proteine nucleari, ad esempio l'assemblaggio della catena respiratoria e la replicazione. Disfunzioni nei fattori nucleari necessari per l'assemblaggio della catena sono stati associati a varie malattie, tra cui la sindrome di Leigh. Alterazioni dell'apparato replicativo del DNA mitocondriale possono essere fonti di duplicazioni o delezioni. Disordini nelle funzioni di replicazione e divisione cellulare possono causare neuropatie ottiche.

1.4 Gli studi evoluzionistici

1.4.1 L'orologio molecolare

Il concetto di orologio molecolare, scoperto da Zuckerkandl e Pauling nel 1965, si basa sull'osservazione che, presa una qualunque proteina, il suo tasso intrinseco di evoluzione molecolare è costante nel tempo ed in tutte le linee. Secondo un principio analogo (ma tenendo presente che il tasso di sostituzione nucleotidica è maggiore di quello di sostituzione aminoacidica a causa della degenerazione del codice genetico – più di una tripletta codifica lo stesso aminoacido – alcune mutazioni del DNA risultano in realtà non rilevabili occupandosi di proteine) è stato possibile stabilire un orologio molecolare associato al DNA mitocondriale, partendo dall'analisi di specie di cui erano disponibili dati sui tempi di divergenza. Il tasso

medio di evoluzione risulta pari a 1-2% per milioni di anni (cioè 1-2 mutazioni su cento paia di basi ogni milione di anni). Questo orologio molecolare ha un ticchettio molto veloce utilissimo nello studio di popolazioni che si sono differenziate in tempi recenti, proprio come quelle umane (Wilson et al, 1985; Howell et al, 2004).

Le differenze evolutive possono essere espresse in due modi; si può misurare il tasso di evoluzione (i cambiamenti avvenuti lungo una linea) contando le mutazioni che si sono accumulate a partire dal DNA ancestrale fino al suo attuale discendente oppure si può calcolare il tasso di divergenza (numero di mutazioni tra due linee) tramite l'analisi delle mutazioni verificatesi tra due linee presenti in una popolazione rispetto all'antenato comune ad entrambe. Nel caso del DNA mitocondriale queste mutazioni si sono accumulate seguendo la linea materna ed i polimorfismi che si sono generati durante l'evoluzione possono essere analizzati per studiare il grado di divergenza tra gli individui e la loro distribuzione tra i vari gruppi umani (Howell et al, 2004).

1.4.2 La variabilità del DNA mitocondriale applicata agli studi evoluzionistici di popolazioni umane

Relativamente all'evoluzionee umana, sono stati formulati due modelli principali: il modello **multiregionale** e quello dell'***Out of Africa***. Secondo il primo *l'Homo sapiens* si sarebbe evoluto indipendentemente e contemporaneamente nelle varie aree del mondo, a partire da un comune predecessore africano (probabilmente *Homo ergaster*) fuoriuscito dalla terra di origine circa 2 milioni di anni fa. L'ipotesi è

compatibile con i dati archeologici delle varie aree geografiche, poiché sussiste una certa continuità anatomo – morfologica nelle sequenze di ritrovamenti fossili. Il suo punto debole risiede nella difficoltà della realizzazione di un'evoluzione parallela possibile unicamente grazie ad un flusso genico considerevole e continuo tra le popolazioni. Secondo il modello dell'Out of Africa la suddivisione in realtà rispecchia un'origine africana recente (circa 200000 anni fa) e unica di tutte le altre popolazioni, poiché dei due rami principali in cui sarebbe diviso l'albero del DNA mitocondriale, uno porta esclusivamente tipi africani, mentre l'altro è composto da tutte le linee osservate nel resto del mondo e dai restanti tipi mitocondriali africani (vedere oltre).

Tutti gli studi effettuati prima del 1989 si basarono sull'analisi elettroforetica dei polimorfismi di lunghezza dei frammenti di restrizione (RFLP) mediante southern blotting o marcatura terminale; solo dai primi anni novanta fu introdotta la tecnica di sequenziamento. La prima ricerca fu effettuata da Brown nel 1980, il quale analizzò 21 individui di varie regioni geografiche utilizzando 18 enzimi di restrizione. La quantità di variazione nucleotidica ottenuta, espressa come numero di sostituzioni per coppie di basi analizzate, fu 0.0018. Utilizzando quindi un tasso evolutivo dell'1% per milione di anni fu possibile stimare la data di speciazione dell'*Homo sapiens*, cioè 180000 anni fa. In realtà il dato non era conclusivo, dato il piccolo numero di individui analizzati e il numero di enzimi, ma da quel momento in poi fu chiaro che per ottenere dati rilevanti occorreva concentrare gli studi su singole aree geografiche o popolazioni (Brown, 1980).

Nel 1983 Johnson e collaboratori analizzarono un campione di circa 200 individui (appartenenti alle popolazioni bosciamani, bantu, nativi americani e asiatici) utilizzando 5 enzimi di restrizione ed evidenziando 35 aplotipi, di cui la maggior varietà era concentrata nelle popolazioni africane. Questo risultato portò a concludere che la nostra specie avesse un'origine africana. L'utilizzo dell'analisi di restrizione "a bassa risoluzione" (utilizzando, cioè, dai 5 ai 6 enzimi di restrizione), ha rivelato un albero filogenetico "star-like", cioè "simile ad una stella", con un tipo mitocondriale posto al centro, condiviso da diverse popolazioni del globo.

Utilizzando l'analisi di restrizione "ad alta risoluzione", condotta con 12 enzimi di restrizione su 147 individui provenienti da diverse aree geografiche, Cann et al (1987) ricostruirono un albero filogenetico dell'mtDNA molto più risolto. I dati ottenuti furono analizzati con il metodo della massima parsimonia (sistema che tende a ricostruire l'albero che minimizza il numero di cambiamenti necessari per ottenere l'albero osservato) che ha permesso di concludere che l'Africa è la fonte più verosimile di tutte le linee osservate nel resto del mondo (Relethford, 2008). In questo albero, si è osservato che un set di sequenze Africane risultava essere una radice di altre sequenze, africane e non. Questa ipotesi dell'"Out of Africa" (OOA) e della discendenza dall'africana "Eva mitocondriale" dell'uomo moderno, proposta da Cann et al (1987) è stata di fondamentale importanza nello stabilire il potenziale della molecola dell'mtDNA come marcatore genetico per gli studi sull'evoluzione umana. Inoltre l'origine africana dell'uomo anatomicamente moderno fu datata a circa 200000 anni fa e si ricavò un modello genetico, definito *"Ipotesi del giardino*

dell'Eden" (GOE), confermato peraltro da studi successivi (Vigilant et al, 1991), secondo il quale l'espansione demografica e geografica delle popolazioni attuali avvenne contemporaneamente alla diversificazione degli aplotipi.

Una rianalisi successiva (Templeton 1992) di questi dati riscontrò alberi filogenetici più brevi in lunghezza rispetto all'albero costruito con il metodo della massima parsimonia (MPT), da parte del gruppo di Vigilant. Alcuni degli alberi filogenetici costruiti da Templeton non sostenevano la base africana dell'evoluzione umana (ipotesi debole del GOE, secondo la quale circa 100000 anni fa la popolazione ancestrale si sarebbe suddivisa in molti gruppi parzialmente isolati che successivamente si espansero e si ridussero, accumulando le variazioni attuali prima dell'espansione massiva, non contemporaneamente). Questo portò ad un periodo di intenso dibattito, ma ulteriori dati e metodi perfezionati nella costruzione di alberi filogenetici, hanno alla fine reso solida l'ipotesi dell'OOA, proprio utilizzando le sequenze della regione di controllo (vedasi, per es. Penny et al, 1995; Watson et al, 1997).

Un importante studio su larga scala fu effettuato sui nativi americani e mise in evidenza la loro origine asiatica (Torroni et al, 1993). La nomenclatura base degli aplogruppi fu definita proprio in questo contesto. Successivamente la caratterizzazione di altre popolazioni ha reso possibile la definizione di altri aplogruppi e sotto – aplogruppi (vedi oltre).

L’emergenza della “mito genomica” si è fatta sentire in maniera incisiva con la pubblicazione di 53 genomi umani mitocondriali, completi, appartenenti a differenti popolazioni, nel tardo 2000, da parte di Ingman e colleghi (Ingman et al, 2000; Maca-Meyer et al, 2001). Il robusto albero filogenetico così ricavato ha confermato studi precedenti (Cann et al, 1987; Vigilante t al, 1991), in particolare l’osservazione che gli uomini sono migrati fuori dall’Africa per popolare il mondo. Tra gli studi più importanti sul DNA antico che sostennero l’ipotesi dell’Out of Africa, c’è quello di Pääbo sull’*Homo neanderthalensis*, cominciato nel 1997. Il reperto analizzato fu trovato nella grotta di Feldhofer nel 1856. La sequenza del frammento estratto presentava un numero di mutazioni tre volte maggiore rispetto alle differenze medie che si riscontrano tra le popolazioni umane attuali, indicando un percorso evolutivo diverso (Krings et al, 1997).

Parallelamente, uno degli studi più rilevanti a sostegno di tale ipotesi è stato quello cominciato nel 1999 e terminato nel 2000 da Ovchinnikov, il quale ha analizzato il genoma mitocondriale di un neandertaliano datato 29000 anni fa proveniente dalla grotta Mezmaiskaya nel Caucaso settentrionale (Ovchinnikov et al, 2000). La sequenza si trova sicuramente fuori dalla variabilità dell’umanità attuale, confermando i dati ottenuti dagli antropologi fisici e molecolari che tra *neanderthalensis* e *sapiens* non si ebbe alcun mescolamento genico, ma che i primi rappresentano un ramo “morto” del nostro albero evolutivo. Gli studi sull’ *Homo neanderthalensis* proseguono ancora oggi e finora tutti i dati hanno confermato la teoria (Krings et al, 2000; Schmitz et al, 2002).

Negli anni a seguire, Finnila et al (2000) utilizzarono l'elettroforesi su gel conformazione sensibile (CSGE) per l'analisi delle sequenze delle regioni codificanti di 22 soggetti finlandesi, appartenenti all'aplogruppo U (vedere oltre). Finnila et al (2001) diedero seguito a tali studi, mediante l'analisi filogenetica di 192 genomi mitocondriali di finlandesi, che ricompresero tutti gli aplogruppi dei Caucasici Europei. L'informazione più grande ricavata dai genomi completi ha indicato un "progredire del tempo" ("coming of age") relativo agli alberi medesimi (Richards & Macaulay, 2001).

Man mano che più laboratori producevano dati di genomi mitocondriali completi, i frutti della mito genomica si sono prodotti in un gran numero di settori. Oltre alla conferma dell'ipotesi dell'"Out of Africa" (Ingman et al, 2000; Maca-Meyer et al, 2001), vi è stata anche quella, rigorosa, dell'assenza di ricombinazione nell'mtDNA (Ingman et al, 2000; Elson et al, 2001), a dispetto delle innumerevoli disquisizioni che sostenevano il contrario. Interi genomi mitocondriali sono stati utilizzati per scopire nuovi polimorfismi associati ad aplogruppi (Finnila et al, 2001; Herrnstadt et al, 2002): Torroni et al (2001) hanno osservato differenze nel tasso di mutazione dei sottotipi dell'aplogruppo africano L2. Mishmmar et al (2003) hanno proposto che la selezione giochi un ruolo nella variazione del genoma mitocondriale tra differenti aplogruppi, e Meyer e von Haeseler (2003) hanno studiato i tassi di sostituzione sito-specifica nell'intero genoma mitocondriale.

Studi dettagliati sull'evoluzione molecolare, peraltro, avevano dimostrato già da tempo che il DNA mitocondriale evolve in maniera complessa. Ad esempio,

Aquadro e Greenberg (1983) chiarirono che la composizione in basi non è bilanciata all'interno della molecola, ed hanno verificato che le transizioni hanno una frequenza significativamente superiore rispetto alle transversioni (rapporto di 32:1). Inoltre i medesimi hanno individuato diversi casi di reversioni, o omoplasie, nelle sequenze da loro esaminate, e riscontrato che la distribuzione delle sostituzioni non era casuale.

La variabilità di ciascun sito rispetto ad un altro, relativamente ai tassi di mutazione nella regione di controllo, è una delle più evidenti peculiarità dell'mtDNA. La maggior parte dei siti nella regione di controllo ha un tasso di mutazione molto basso, o non varia assolutamente, mentre altri siti mutano in maniera estremamente veloce. Questo enorme tasso di eterogeneità è stato riconosciuto da numerosi studi, utilizzando un'ampia varietà di metodi (Wakeley, 1993; Hasegawa et al, 1993; Meyer et al, 1999; Excoffier e Yang, 1999; Pesole e Saccone, 2001; Malyarchuck et al, 2002). Ciononostante, la complessità dell'evoluzione dell'mtDNA deve ancora essere caratterizzata e spiegata in maniera completa. Ad esempio, studi recenti su discendenze (pedigree) hanno determinato un inaspettato alto ed empirico tasso di mutazione nella regione codificante (Howell et al, 1996; Parsons et al, 1997; Sigurdardottir et al, 2000; Heyer et al, 2001). Le stime iniziali dei tassi di mutazione, determinati dagli studi sui pedigree sono risultate essere di circa dieci volte superiori a quelle ricavate dagli studi filogenetici (Howell et al, 1996; Parsons et al, 1997). Pertanto si è dibattuto a lungo se i siti ad alta velocità di mutazione fossero la causa di questa discrepanza (Paabo, 1996; Jazin et al, 1998; Parsons e Holland, 1998). Howell et al (2003) hanno pubblicato i risultati di uno studio su larga scala di pedigree, ed

hanno riconfermato la discrepanza in oggetto. E' quindi lecito pensare che non vi sia un unico fattore che rende conto di questa, ma che punti caldi di ricombinazione ("Mutational hotsposts"), la deriva genetica, la selezione, e la non totale capacità dei metodi filogenetici di rendere conto in maniera esatta dell'eterogeneità del tasso di mutazione, possano contribuire alla disparità osservata tra i tassi di mutazione ricavati dagli studi filogenetici e quelli ricavati empiricamente dai pedigree.

L'eterogeneità nel tasso di mutazione può simulare l'espansione della popolazione (Lundstrom et al, 1992; Aris-Brosou e Excoffier, 1996) ed influire sulle stime dei parametri genetici di popolazione, come quella relativa all'effettiva dimensione della popolazione (Fu, 1994; Cavalli-Sforza et al, 2003) ed alla calibrazione dell'orologio molecolare (Adachi e Hasegawa, 1995; Yang, 1996; Howell et al, 2004). Wakeley (1994) ha dimostrato che il non tener conto della variazione del tasso di mutazione può portare ad un'enorme sottostima del rapporto transizione : transversione.

1.5 Gli aplogruppi e la storia dell'uomo

L'uscita della nostra specie dal continente africano risale a circa 70000 anni fa, quando alcuni *sapiens* hanno abbandonato il Corno d'Africa per dirigersi verso le coste dell'Arabia, dell'Iran, dell'India, per arrivare in Asia orientale (Maca-Meyer et al, 2001). Da lì, l'espansione continuò fino all'Asia meridionale e circa 40000 anni fa raggiunse Australia e Nuova Guinea (fig. 1.9). La migrazione verso l'Europa sembra

essere iniziata solo 35000 anni fa, probabilmente a causa delle condizioni climatiche sfavorevoli del nostro continente. L'ultima migrazione preistorica fu quella verso le Americhe, a partire dall'Asia settentrionale, attraverso il corridoio della Beringia, avvenuta in tre ondate successive (nel periodo compreso tra 30000 e 6000 anni fa).

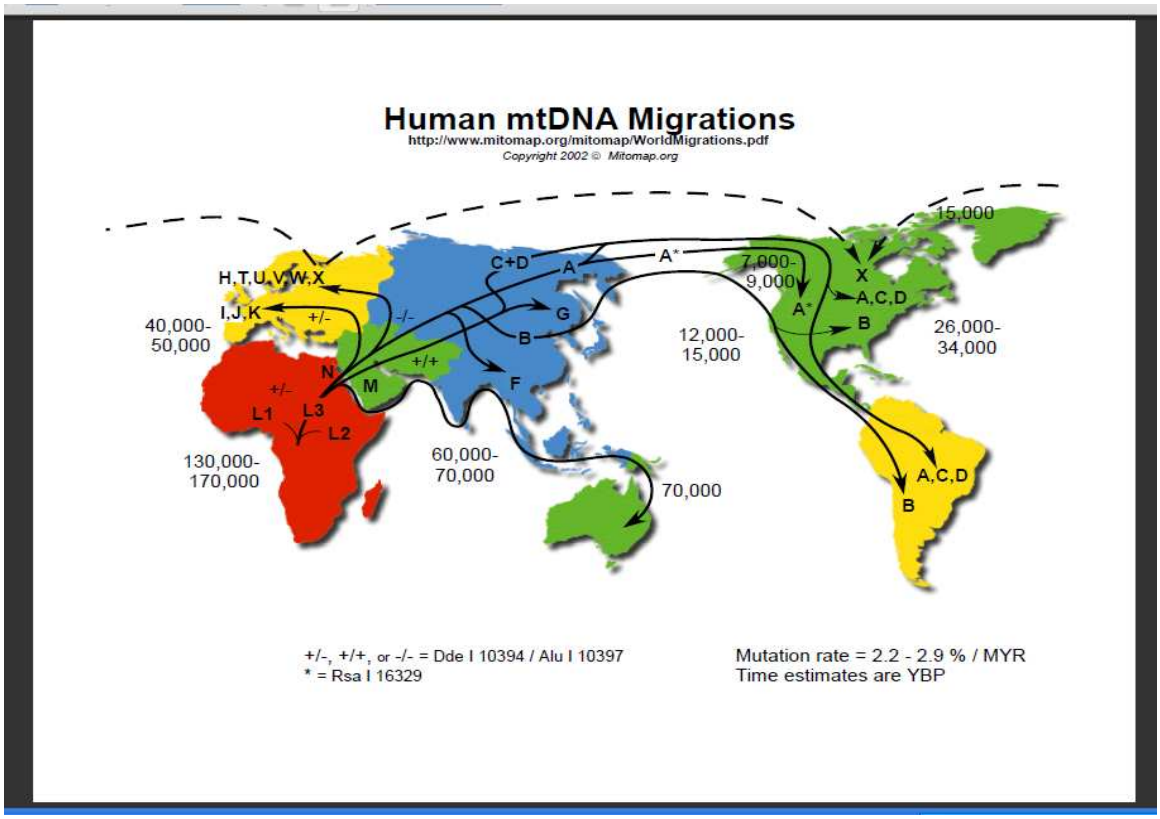


Fig. 1.9 I flussi migratori della popolazione umana descritti dal DNA mitcondriale.

Recentemente (Olivieri et al, 2006) è stata dimostrata una migrazione a partire dall'Asia sud occidentale verso l'Europa e il Nord Africa, avvenuta circa 40000 anni fa, durante un periodo inter – glaciale, su rotte diverse da quelle seguite durante le prime espansioni verso l'oriente.

1.5.1 Gli aplogruppi principali

I polimorfismi del mtDNA sono associati a ed in particolari aplotipi, i quali a loro volta possono essere raggruppati in aplogruppi (un aplogruppo, cioè, risulta costituito da un gruppo di aplotipi derivati da uno stesso mtDNA ancestrale e determinato da una mutazione specifica) (Torroni et al, 1993). Gli studi condotti su popolazioni di tutto il mondo portano alla conclusione che la variabilità del mtDNA è correlata all'origine etnica e geografica di un individuo.

L'analisi della variazione degli RFLP nelle varie popolazioni, prese nella loro globalità, ha consentito l'identificazione di un certo numero di rami monofiletici, nei quali è sato possibile classificare tutti i DNA mitocondriali. Questi "clades", o "aplogruppi", sono stati differenziati grazie a mutazioni "antiche", verificatesi migliaia di anni fa. Marcatori continente-specifici sono stati individuati per gli Africani (Chen et al, 1995), i Caucasici di origine europea (Torroni et al, 1994; 1996) e gli Asiatici (Schurr et al, 1990). Gli aplogruppi vengono designati, in letteratura, da una lettera maiuscola seguita da un numero, che rappresenta il sottogruppo ("sub cluster") dell'aplogruppo (Schurr et al, 1990; Torroni et al, 1996; Richards et al, 1998). Quasi tutti gli mtDNA africani possono essere classificati in uno dei tre aplogruppi principali (L1, L2 o L3). L'aplogruppo L1 è ritenuto essere quello ancestrale, dal momento che lo si ritrova alla radice dell'albero filogenetico. I super-aplogruppi Euroasiatici M ed N sono ritenuti aver lasciato l'Africa circa 50.000-

60.000 anni fa, per poi dare origine agli aplogruppi attuali in Asia (A, B, C, D, F) ed in Europa (H, I, J, K, T, U, V, W, X) (Tab. I.4).

Dal momento che il genoma mitocondriale è una unica molecola in linkage, i motivi di sequenza nella regione di controllo correlano fortemente con gli aplogruppi definiti mediante l'analisi in RFLP (Torroni et al, 1996; Macaulay et al, 1999a). Ad esempio, il polimorfismo 73A è tipicamente associato all'aplogruppo H (Torroni et al, 1996; Macaulay et al, 1999a; Allard et al, 2002), rendendo questa posizione un'omoplasia di non raro riscontro nell'analisi filogenetica. Esistono tuttavia diversi esempi in letteratura, nei quali individui che appartengono all'aplogruppo H hanno subito la reversione a 73G (Torroni et al, 1996; Macaulay et al, 1999a; Allard et al, 2002); dal momento che questo motivo è normalmente associato all'aplogruppo U, è impossibile classificare in maniera univoca gli individui che possiedono il polimorfismo 73G come appartenenti ad un dato aplogruppo, senza possedere ulteriori informazioni a livello della regione codificante.

In Africa l'aplogruppo principale è L (Chen et al, 1995) che, con i suoi sotto – aplogruppi (L1a, L1b, L2 ed L3), rappresenta quasi la totalità della popolazione sub – sahariana. L'aplogruppo più diffuso è L1 (circa 50%), seguito da L2 e L3. Quest'ultimo rappresenta una piccola frazione della popolazione, ma sembra essere il progenitore di ben la metà delle linee di mtDNA in Europa, Asia e Nativi Americani. L'origine dell'aplogruppo L è datata tra 162000 e 123000 anni fa, considerando il tasso di evoluzione 2.2-2.9% per milione di anni (Torroni et al, 1994).

CONTINENTE	APLOGRUPPI PRINCIPALE	APLOGRUPPI MINORI	% aplog. principali
Africa	**L**	**M, U**	**76%**
Europa	**H, I, J, K**	**M, T, U, V, W, X**	**64%**
Asia	**M**	**A, B, F**	**55%**
Nat. Americani	**A**	**B, M**	**100%**

Tab. I.4 Schema sintetico degli aplogruppi presenti nel Mondo

In Asia l'aplogruppo M si riscontra in più della metà degli individui (circa 55%), ed è suddiviso nei sotto – aplogruppi C, D, G, E. La maggior parte dei restanti individui è compresa negli aplogruppi minori A, B e F.

Nei nativi americani sono stati trovati solo 4 aplogruppi asiatici: A, B, C e D. Ciò ha permesso di retrodatare la colonizzazione delle Americhe. L'aplogruppo principale risulta essere l'A, uno dei più antichi (Torroni et al, 1993).

Per quanto riguarda l'Europa, l'analisi è più complessa (vedi oltre). L'aplogruppo principale è l'H, che raggruppa quasi il 40% della popolazione; sono stati definiti, inoltre, altri sottoaplogruppi, che mostrano una caratteristica distribuzione regionale. Tra questi, H1 ed H3 sono di comune riscontro nell'Europa occidentale, essendosi espansi dal rifugio Franco-cantabrico (Achilli et al, 2004; Loogväli et al, 2004), mentre un sottotipo di H2, definito dalla transizione in 951, è tipico dell'Europa orientale e dell'Asia. Sono stati evidenziati altri 9 aplogruppi (tra cui anche M), di cui I, J e K sono i più rappresentati (Torroni et al, 1996).

1.5.2 Aplogruppi in Europa e in Italia

1.5.2.1 Geografia dell'Europa e dell'Italia

L'Europa è una penisola dell'Asia di dimensioni relativamente piccole (circa 10 milioni di chilometri quadrati, pari a 1/14 delle terre emerse) rispetto alla densità di popolazione (inferiore solo rispetto a quella asiatica). E' separata dall'Asia a Nord dal Mar Glaciale Artico, a Est dal fiume Ural e a Sud dal Mar Mediterraneo. A Ovest confina con l'Oceano Atlantico. L'Europa è caratterizzata da un clima suddivisibile in quattro regioni principali: sub-tropicale, oceanico, continentale e sub-artico. Ciò ha causato una distribuzione non omogenea degli abitanti, che nel corso degli anni si sono concentrati maggiormente in zone temperate e fortemente industrializzate. Una delle zone più popolate è proprio la penisola italiana, il cui clima mediterraneo ha costituito, soprattutto dopo l'ultima glaciazione, un ambiente accogliente e ricco di risorse, adatto all'agricoltura.

1.5.2.2 Storia archeologica dell'Europa

Durante il Paleolitico (periodo che va da circa 2 milioni di anni fa a circa 100000 anni fa, termine dell'ultima glaciazione), il popolamento dell'Europa è attribuito *all'Homo heidelbergensis* (dal nome della località del primo ritrovamento, nei pressi di Mauer, vicino al villaggio di Heidelberg, in Germania). Sembra che l'uscita dall'Africa di questo ominide risalga a circa 1 milione di anni fa, anche se nella Europa dell'est sono stati ritrovati manufatti più antichi di circa 500000 anni. La

datazione di un milione di anni comunque, sembra quella più accettabile per quanto riguarda un occupazione stabile dei territori europei. In base ai ritrovamenti fossili, *Homo heidelbergensis* probabilmente fu un discendente dell'*Homo ergaster*. Le produzioni litiche corrispondenti sono principalmente quelle dell'Acheuleano, con evoluzione differente a seconda delle zone geografiche per quanto riguarda la forma e le tecniche di scheggiatura delle bifacciali, e il Clactoniano, diffuso soprattutto nelle aree mediterranee.

Tra i 100000 e i 40000 anni fa, ad un periodo climatico particolarmente rigido (prima glaciazione würmiana), fece seguito l'espansione europea dell'*Homo neanderthalensis* e la scomparsa graduale dell' *heidelbergensis* al quale viene fatto riferimento per le culture del Musteriano (di cui fanno parte una serie di innovazioni culturali, come il rito funerario, nuove tecniche di scheggiatura e strategie economiche). Circa 40000 anni fa l'*Homo sapiens* colonizzò l'intero continente europeo. L'*Homo sapiens* è un uomo anatomicamente moderno, che sostituì completamente in soli 10000 anni il cugino non imparentato *Homo neanderthalensis*. I primi ritrovamenti di *sapiens* risalgono al 1866 a Solutrè e al 1868 ad Abri Cro – Magnon (Francia). La prima cultura dell'uomo moderno è l'Aurignaziano (dal sito francese di Aurignac) portatrice di strategie tecniche vincenti, come la formazione di accampamenti organizzati e costruzione di strutture coperte, la lavorazione dell'osso e del corno, la creazione di oggetti ornamentali e la nascita dell'arte figurativa. Successivamente si susseguirono una serie di culture, la Gravettiana (tra 29000 e 20000 anni fa, forse a causa di una migrazione dall'oriente), la Solutreana (tra 20000

e 17000 anni fa, soprattutto in Europa occidentale), la Maddeleniana (fino a 12000 anni fa, riguardante tutta l'Europa) e l'Aziliana (sorta in Europa occidentale circa 10000 anni fa).

A partire dal Neolitico cominciarono una serie di cambiamenti sostanziali. Il primo cambiamento evidente fu un aumento considerevole della popolazione, concomitante alla nascita dell'agricoltura (orzo, grano), proveniente dal Medio Oriente, e dell'allevamento (ovini, suini, bovini). L'uomo passò dalla vita nomade a quella sedentaria. Altri cambiamenti riguardarono le tecniche di lavorazione della pietra e l'uso della ceramica. Nel 1973 Ammerman e Cavalli – Sforza proposero un modello per descrivere il processo dell'aumento demografico e della conseguente espansione della cultura agricola; il modello, detto dell'onda di avanzamento, prevedeva migrazioni locali caratteristiche di popolazioni rurali dalle abitudini semi – nomadi, che si basavano su un sistema "taglia – brucia". Questi movimenti verso zone di confine e di successiva espansione demografica avrebbero così creato un fronte d'onda, a partire dalla Grecia verso i Balcani. Un altro tipo di ceramica, caratterizzato da un disegno lineare, è il marcatore archeologico dell'espansione che a partire dal Danubio centrale giunse fino al Mare del Nord. Tra le ultime zone ad essere occupate sono le Isole britanniche e la Scandinavia, a causa del loro clima particolarmente inospitale.

Tra il 4000 e il 3000 a.C. ebbe inzio l'età del Rame, in cui si collocano le culture dei monumenti megalitici (Dolmen, Menhir), costruzioni che a volte erano

molto elaborate (come nel caso di Stonehenge, vicino a Salisbury). E' probabile che la funzione originaria dei megaliti fosse legata ai riti funerari oppure alle osservazioni astronomiche. Durante l'età del rame, soprattutto nel Nord Europa prese piede la tecnica del nomadismo pastorale e anche l'addestramento dei cavalli sembra collocarsi in questo periodo storico.

L'età del Bronzo si inserisce nella fascia temporale tra il 2000 a.C. e il 1000 a.C., almeno per quanto riguarda l'Europa continentale. In Grecia, probabilmente, l'età del Bronzo comincia antecedentemente, intorno al 3000 a.C.. Durante questo periodo si diffonde la cultura della cremazione delle salme.

Durante il primo millennio avanti Cristo si diffuse la cultura dell'età del Ferro, che sembra accomunare le popolazioni dal punto di vista linguistico piuttosto che da quello genetico (come nel caso dei Celti).

1.5.2.3 Storia archeologica dell'Italia

Il più antico popolamento dell'Italia rientra nelle dinamiche dei movimenti dei gruppi umani che hanno colonizzato stabilmente l'Europa circa un milione di anni fa. Riguardo ai primi insediamenti, i più antichi ritrovamenti sono quelli di Ca' Belvedere di Monte Poggiolo, presso Forlì, seguiti da quelli laziali di Colle Marino e di Arce e Fontana Liri. Sono presenti tutte e due le culture contemporanee, Acheuleana e Clactoniana. La penisola italiana ha accolto gruppi di neanderthaliani

che però a differenza dei predecessori non si spinsero fino alle isole, creando gruppi circoscritti e poco mobili.

Successivamente tutta l'Italia, a partire dalle Prealpi venete fino ad arrivare in Sicilia, fu colonizzata da *Homo sapiens* e dalla cultura Aurignaziana. La successiva cultura Gravettiana occupò solo la penisola, senza spingersi nell'Italia insulare.

Circa 24000 anni fa l'ampliamento dei ghiacciai dell'arco alpino determinò la formazione di una barriera culturale tra i gruppi peninsulari ed il resto dell'Europa. Alla fine di questo processo di isolamento, l'Italia e le regioni mediterranee vicine cominciarono un processo individuale di evoluzione culturale. Nacque così l'Epigravettiano, che si protrarrà fino a 10000 anni fa circa, data in cui convenzionalmente si fa iniziare l'Olocene, l'era geologica in cui viviamo.

Anche in Italia, come nel resto dell'Europa, il Neolitico è caratterizzato dall'espandersi della cultura agricola; in particolare, tutte le zone mediterranee, a causa della facilità di spostamenti lungo le coste e del clima favorevole, furono colonizzate molto rapidamente. In Italia, nella Francia meridionale e nella Spagna orientale emerse un andamento culturale locale, che si distinse da quelli del resto d'Europa per il tipo di vasellame, detto "ceramica impressa o cardiale".

In Italia sono presenti alcuni tra i più antichi reperti di monumenti megalitici dell'età del rame, soprattutto in Sardegna, Sicilia e Puglia. Le culture dell'età del Bronzo, al contrario, hanno lasciato tracce al Centro e al Nord del Paese.

Durante l'età del Ferro, l'Italia vide sorgere una delle sue culture più significative, la civiltà etrusca, caratterizzata da una lingua non – indoeuropea. Negli anni successivi tutte le coste mediterranee furono colonizzate dalla civiltà greca, ad eccezione della Sardegna, della parte occidentale della Sicilia, delle coste della penisola iberica e dell'Africa del Nord, in cui predominò quella fenicia, seguita da quella cartaginese. Successivamente tutte queste zone nonché gran parte dell'Europa continentale caddero sotto la dominazione romana. Il resto è "storia".

1.5.2.4 Genetica dei polimorfismi classici e flussi migratori

L'Europa è il continente che si distingue per il più alto contenuto di informazioni genetiche, ma è anche quello più difficile da interpretare filogeneticamente. Infatti in Europa molte popolazioni che sono divise dall'appartenenza ai vari paesi, in realtà sono eterogenee e formate da vari gruppi genetici e linguistici. Attualmente la ricerca sul pool genico sembra evidenziare una formazione recente, avvenuta in due fasi principali, la prima iniziata circa 40000 anni fa e la seconda dovuta all'espansione del Neolitico dai gruppi agricoli provenienti dall'oriente. I filoni principali al riguardo sono due, uno (teoria della diffusione culturale) che considera l'attuale popolazione europea derivante principalmente dalle espansioni paleolitiche, l'altra, sostenuta da Cavalli – Sforza (teoria della diffusione demica) che propone un'origine genetica orientale risalente all'espansione degli agricoltori nel Neolitico. Lo studio dei polimorfismi classici (gruppi sanguigni e antigeni HLA) rivela che il continente europeo è geneticamente omogeneo con

l’eccezione di alcune popolazioni, i Sardi, gli Islandesi, i Baschi e i Lapponi (la popolazione più lontana dalle altre).

Per quanto riguarda l’Italia, nel 1988 Piazza e collaboratori hanno pubblicato un’analisi delle componenti principali basata su 34 frequenze geniche e hanno rappresentato le mappe sintetiche delle prime tre componenti, dimostrando che la Sardegna non è geneticamente correlata alla penisola. Di conseguenza la regione è stata esclusa dalle analisi. Dal punto di vista genetico l’Italia settentrionale si rivela simile all’Europa centrale, mentre le regioni centrali e meridionali hanno più punti in comune con la Grecia e gli altri paesi mediterranei, dato confermato dalle ricerche archeologiche (vedi sopra). Nel complesso, i valori ottenuti pongono l’Italia in un sottogruppo europeo insieme a Spagna e Portogallo.

1.5.2.5 Il DNA mitocondriale europeo

Durante gli anni ’90, lo studio del DNA mitocondriale ha portato all’identificazione di aplogruppi specifici per l’Europa (Torroni et al. 1994, 1996, 2000, 2001, 2006). Il primo studio del 1994 fu effettuato su 175 DNA caucasici provenienti dagli Stati Uniti e dal Canada, tramite PCR di amplificazione e successiva analisi dei polimorfismi di restrizione. L’analisi dimostrò che tutti i campioni erano ascrivibili per il 59% in 4 aplogruppi, H, I, J e K di cui l’H da solo costituiva il 40%.

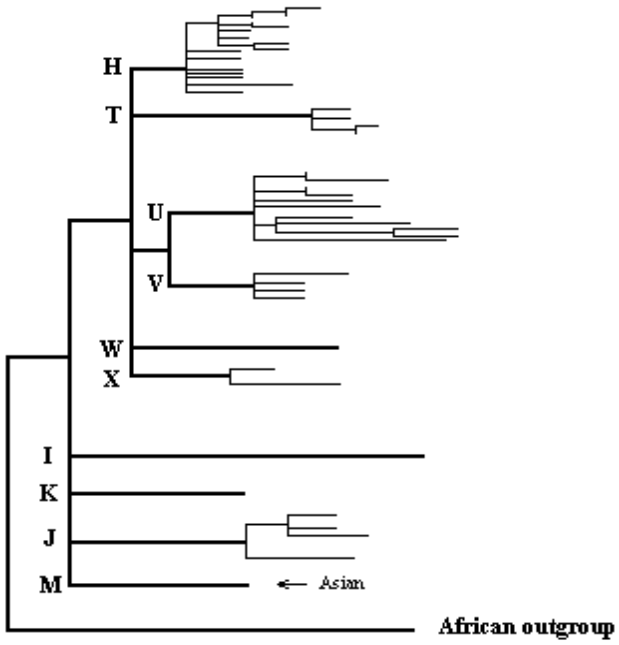


Fig. 1.10 Albero filogenetico degli aplogruppi europei. Albero ottenuto col metodo della massima parsimonia rappresentante i risultati ottenuti nel lavoro di Torroni del 1996 relativamente alle popolazioni Finlandese e Svedese. L'outgroup utilizzato per la radice è l'aplotipo AF71, aplogruppo L, proveniente dal Senegal (Chen, 1995).

Nel 1996 l'analisi di tre popolazioni europee (Svedese, Finlandese e Toscana, scelta sulla base di analisi linguistiche effettuate da Cavalli – Sforza come rappresentative di gruppi principali europei) permise un esame più completo di questi aplogruppi. Furono evidenziati 10 aplogruppi, tra cui l'aplogruppo M, tipico dell'Asia, e i quattro già identificati precedentemente. A questi si affiancarono gli aplogruppi T, U, V, X e W. I dati confermarono che l'aplogruppo H è il più rappresentato (fig. 1.10 e tab. I.5). Tramite il paragone con i dati dell'epoca, fu stabilito che gli aplogruppi H e K erano principalmente, ma non esclusivamente, europei (di ognuno ne erano stati osservati tre in soggetti non – europei); che l'aplogruppo J era esclusivo dell'Europa; che gli aplogruppi T, I e W non potevano

essere collocati tra gli aplogruppi continente – specifico, poiché non c'erano ancora dati riguardo l'enzima di restrizione BstNI (identificativo dei suddetti aplogruppi) nei campioni non – europei analizzati; che l'aplogruppo X probabilmente era di derivazione asiatica, ma difficilmente analizzabile a causa del sito di restrizione al nucleotide 1715, perso varie volte in maniera indipendente nel corso della storia; che l'aplogruppo U era tipico dell'Europa e presente in alcuni individui dell'Africa Sub – Sahariana; che l'aplogruppo V non era classificabile con certezza per mancanza di informazioni. In conclusione, gli aplogruppi H, I, J, K, T e W sembravano essere confinati essenzialmente in Europa ed essere il frutto di un'evoluzione avvenuta dopo l'arrivo in Europa dei primi *sapiens*, mentre per l'aplogruppo U era ipotizzabile un'origine africana più diretta.

Ad oggi, dopo altri studi e con più risultati a livello mondiale, si possono ritenere ancora validi i risultati dei lavori appena citati, anche se l'albero filogenetico risulta molto più accurato e le relazioni filogenetiche più complesse grazie al sequenziamento totale del DNA mitocondriale, che ha permesso la determinazione di numerosi marcatori polimorfici (Finnilä et al, 2001; Kivisild et al, 2006). Le frequenze per gli aplogruppi europei sono sintetizzate nella tabella I.5.

APLOGRUPPI	FREQUENZE
H	**40.9%**
T	**12.7%**
U	**14.3%**
V	**3.2%**
X	**4.1%**
W	**2.1%**
I	**2.1%**
J	**10.5%**
K	**8.0%**
Altri	**2.1%**

Tab. I.5 Frequenze degli aplogruppi europei (Torroni et al, 1996)

L'età della colonizzazione europea si può calcolare in base alla divergenza tra le biforcazioni dell'albero filogenetico, posta tra i 51000 e i 39000 anni fa. L'aplogruppo U, come già accennato, probabilmente rappresenta una delle linee mitocondriali fondatrici europee.

L'aplogruppo H sembra essersi originato nell'Europa dell'est circa 30000 – 25000 anni fa. Si sarebbe poi espanso nel resto del continente prima della seconda pleniglaciazione, contemporaneamente all'espansione della cultura Gravettiana (25000 – 20000 anni fa). La tabella I.6 indica la frequenza dell'aplogruppo H in alcuni paesi e aree geografiche. Tra i sotto – aplogruppi in cui può essere suddiviso, almeno per quanto riguarda l'Europa dell'Ovest, i due principali sono H1 e H3 (Achilli et al, 2004). Infatti recenti studi di Roostelu e collaboratori hanno evidenziato la presenza di numerosi e frequenti nuovi sotto-aplogruppi nelle regioni dei balcani e del Medio Oriente (Roostelu et al, 2007).

Regione	Frequenza H (%)
Senegal	-
Marocco	36.8
Algeria	25.6
Egitto	1.4
Spagna	39.4
Francia	47.2
Olanda	38.2
Austria	44.8
Albania	48.6
Ungheria	42.3
Polonia	37.2
Turchia	26.0
Iraq	19.9
Penisola Arabica	10.6
Asia centrale	11.2

Tab. I.6 Distribuzione dell'aplogruppo H nelle popolazioni

L'aplogruppo V probabilmente si originò in Africa del Nord circa 12500 anni fa ed è molto diffuso, oltre che nell'area originaria, nell'Europa nord – occidentale.

Per quanto riguarda gli aplogruppi H, I, X e K secondo gli studi condotti da Torroni e collaboratori, questi risalirebbero a circa 20000 anni fa.

L'aplogruppo J sembra essere più recente (10000 anni fa, primi anni del Neolitico) e di origine mediorientale.

L'aplogruppo più antico è U, in particolare il suo sottoaplogruppo U5, probabilmente giunto in Europa nel periodo Aurignaziano, circa 30000 anni fa, portato da colonizzatori provenienti dall'est.

Le altre sequenze hanno un'origine recente, in parte africana e in parte asiatica.

Tutti le ricerche che hanno portato a questi risultati confermano che la maggior parte del pool genetico degli attuali europei non deriva dall'espansione neolitica, ma era già presente durante il Paleolitico. In particolare, sembra che il rifugio Franco – Cantabrico dell'ultima glaciazione, situato a cavallo dei Pirenei, sia stato uno dei pochi a garantire la sopravvivenza, rappresentando così un collo di bottiglia; i risultati ottenuti al riguardo hanno dimostrato che gran parte della variabilità genetica presente oggi in Europa proviene proprio da quella zona (Achilli et al, 2004). Nel 2000, Richards e collaboratori hanno ipotizzato che la componente neolitica rappresenti meno di un quarto degli attuali mtDNA europei. Tuttavia è certo che entrambe le espansioni abbiano lasciato eredità rintracciabili a livello genetico.

1.5.2.6 Il DNA mitocondriale della popolazione italiana

Gli studi delle variazioni polimorfiche del DNA mitocondriale presenti in letteratura fino ad oggi non hanno interessato l'intera popolazione italiana. Sono stati studiati i polimorfismi del DNA mitocondriale in alcune zone della Toscana (studi mirati a scoprire l'origine genetica della popolazione etrusca; Vernesi et al, 2004; Achilli et al, 2007), in alcune località del Molise (studi su popolazioni linguisticamente isolate: Babalini et al, 2005) e della Sardegna (gruppi geneticamente isolati: Fraumene et al, 2006). In alcune zone della penisola gli aplogruppi sono stati analizzati per individuare eventuali associazioni con patologie (Ghezzi et al, 2005) e con la longevità (Dato et al, 2004). Studi più estesi hanno coinvolto solo individui

appartenenti all'aplogruppo H (Achilli et al, 2004), per i quali è stato poi possibile ricavare i dati sulla frequenza degli aplogruppi in Italia, relativi all'aplogruppo H ed ai sottoaplogruppi H1 e H3 (tab. I.7).

Area geografica	Frequenza Aplogruppo H (%)	Frequenza sottogruppi	
		H1	H3
Nord Italia	**46.9%**	**11.5%**	**5.0%**
Centro Italia	**35.6%**	**6.3%**	**3.8%**
Sud Italia	**37.9%**	**8.7%**	**2.4%**
Sicilia	**48.9%**	**10.0%**	**2.2%**
Sardegna	**42.5%**	**17.9%**	**8.5%**

Tab. I.7 Frequenza dell'aplogruppo H e dei suoi sottoaplogruppi H1 e H3.

1.6 L'mtDNA nel forense: utilità e vantaggi

Gli studi sul DNA mitocondriale hanno trovato un ruolo importante nella comunità forense del DNA (revisione di Holland e Parsons, 1999); la totale ereditarietà per via materna, con trasmissione alla discendenza a seguito dell'espansione dei mitocondri nell'oocita (Giles et al, 1980), la già ricordata assenza di ricombinazione (Ingman et al, 2000; Elson et al, 2001), cosa che implica che le differenze accumulatesi tra qualsivoglia due molecole, riflettano variazioni subentrate in tempi successivi (dal momento che esse condividono un antenato comune), il tasso di mutazione fino a dieci volte superiore a quello dei geni del DNA nucleare, in singola copia (Brown et al, 1979; Horai et al, 1995), lo hanno reso una molecola

ampiamente studiata e sfruttata nelle discipline forensi. Non da ultimo va ricordato l'elevato numero di copie di tale molecola all'interno della cellula: i loci autosomici nucleari utilizzati nel forense sono presenti solo in duplice copia, in ciascuna cellula, mentre il DNA mitocondriale è presente in circa 500-2000 copie per cellula di mammifero (Piko et al,1976; Michael set al, 1982). E' corrispondentemente più probabile che alcune copie amplificabili dell'mtDNA sopravvivano in campioni altamente degradati, consentendo l'analisi delle tracce, che altrimenti non fornirebbero alcun profilo del DNA nucleare. I campioni tipicamente sottoposti all'analisi per mtDNA comprendono macchie di sangue degradate, ossa, saliva, unghie, e fusti di capelli. La tipizzazione dell'mtDNA per questi ultimi è un'applicazione particolarmente importante, dal momento che capelli caduti sono di comune riscontro quali fonte di prova. Le caratteristiche morfologiche rendono le analisi soggettive, essendo basate sull'esperienza ed il giudizio dell'esaminatore (Bisbing, 1982). In un tentativo di ridurre la soggettività associata all'analisi del capello, un numero crescente di laboratori si sta orientando verso l'analisi del DNA mitocondriale. Inoltre, il DNA mitocondriale può ulteriormente discriminare tra capelli che non possono altrimenti essere esclusi come provenienti dallo stesso individuo.

1.6.1 L'analisi del DNA mitocondriale nel forense

Attualmente, la maggior parte dei laboratori forensi che esegue l'analisi del DNA mitocondriale, si focalizza sull'informazione che deriva dalla sequenza di HVI

ed HVII (Holland et al, 1993; Wilson et al, 1993). Come spesso accade nell'identificazione di resti umani, l'analisi paragona le sequenze relative a tali resti con quelle di un familiare di riferimento, appartenente alla linea materna. Se tali sequenze differiscono in più posizioni, può essere dichiarata una esclusione certa. Se le sequenze sono paragonabili, cioè hanno dei "match", ciò è consistente con l'ipotesi di parentela in linea materna. Come conseguenza della rigorosa ereditarietà materna dell'mtDNA, ci si attende che un individuo abbia il profilo coincidente con quello di tutti i suoi parenti in linea materna ("barring mutation") dei quali ci potrebbe essere un esteso numero nella popolazione. E' anche possibile che individui non imparentati in linea materna possano, per caso, esibire lo stesso assetto HVI/HVII a causa di mutazioni omoplasmiche. Infine, al contrario dei loci STR, il genoma mitocondriale è una molecola legata, non soggetta a ricombinazione. Il prodotto delle frequenze dei polimorfismi in HVI ed HVII non può essere utilizato per determinare la frequenza di un profilo di DNA mitocondriale. Per i motivi sopra elencati, un match di mtDNA di per se stesso non può essere considerato un identificativo conclusivo. Al fine di determinare la significatività di un match dell'mtDNA, si deve fare riferimento alla frequenza osservata di quel polimorfismo in un database rilevante (Holland e Parsons, 1999).

1.6.2 La distribuzione dei Caucasici in base ai dati del DNA mitocondriale

Un preliminare studio internazionale ha portato alla costituzione di un primo database forense, per l'mtDNA, che conteneva oltre 4800 profili. Questo database è

tuttora disponibile via internet (http://www.fbi.gov – vedere Monson et al, 2002). Il range delle sequenze di HVI ed HVII, in questo database, è compreso tra le posizioni 16024 – 16365 per l'HVI e 73-340 per l'HVII. I confronti a coppia ("pairwise") delle sequenze dei Caucasici nel database (N=1655) forniscono un quadro significativo della distribuzione dei tipi HVI/HVII. In questi confronti non si è tenuto conto delle inserzioni nella regione altamente variabile del C-stretch dell'HVII, data l'instabile permutabilità in questa regione (Stewart et al, 2001). La distribuzione di frequenze degli aplotipi nella popolazione Caucasica ha una classica forma ad "L"), continua, ad un estremo della quale si trovano quelle sequenze HVI/HVII che rappresentano tipi "unici", visti, cioè, solo una volta nel database (839/1665 [50.4%] dei tipi HVI/HVII). All'altro estremo della distribuzione vi è un piccolo numero di tipi HVI/HVII comuni e condivisi nel database relativo ai Caucasici. Il tipo HVI/HVII più comune si verifica ad una frequenza di circa il 7% della popolazione (fig. – vedere anche Lutz-Bonengel et al, 2003). Sebbene questa distribuzione sia specifica per i Caucasici, le sue caratteristiche generali si riscontrano pure in tutte le popolazioni studiate finora: la presenza di un gran numero di sequenze, che risultano essere uniche nel database, e di un piccolo numero di tipi comuni, quest'ultimo rappresentante una proporzione apprezzabile degli individui nella popolazione. La distribuzione di frequenza dei tipi HVI/HVII illustra la più grande limitazione dell'analisi del DNA mitocondriale: il piccolo numero di tipi comuni nella popolazione per i quali il potere di discriminazione ("Discrimination Power" o "DP") è basso. Altri database presenti allo stato attuale sono quello dell'FBI, oggi

denominato “CODIS”, e quello europeo, che si sta consolidando e che, come i precedenti, punta allo sviluppo di database di popolazione, di alta qualità, per applicazioni forensi e non: è l’European DNA Profiling Group mitochondrial DNA population database project (EMPOP: http://www.empop.org).

1.6.3 Implicazioni analitiche

Per lo scienziato forense, le sfide associate con la generazione di un profilo, a partire da resti scheletrici altamente degradati rimangono consistenti. Spesso il ricercatore deve spendere un’enorme quantità dello scarso estratto di DNA per genrare una sequenza HVI ed HVII mediante set di piccoli primer che si sovrappongono parzialmente e che hanno come bersaglio brevi ampliconi (150 paia di basi o meno; Gabriel et al, 2001). Circa la metà delle volte, il profilo HVI/HVII che ne risulta sarà osservato come unico nel database di popolazione. Circa il 20% delle volte esso apparterrà ad uno dei tipi comuni (quelli che si riscontrano ad una prevalenza dello 0.5% o più nella distribuzione di frequenza): in questo modo, circa un quinto delle volte lo scienziato forense si trova davanti a risultati per i quali il potere di discriminazione è particolarmente basso.

Capitolo 2

SCOPO DELLA TESI

2.1 Sequenziamento totale del DNA mitocondriale

Scopo principale di questo studio è restringere una delle attuali limitazioni dell'analisi del DNA mitocondriale nel forense: la non capacità di distinguere tra individui non correlati che condividono tipi comuni HVI/HVII. Si può ipotizzare che accedendo all'informazione sulle sequenze, relativa alla regione codificante del genoma mitocondriale, si può ricavare un'ulteriore discriminazione dei tipi comuni HVI/HVII. A tale scopo, si è deciso di sequenziare gli interi genomi mitocondriali di individui non correlati - includendo l'intera regione di controllo per la prima analisi discriminativa - al fine di individuare i polimorfismi che caratterizzano la popolazione in esame, la loro distribuzione rispetto a quella di riferimento e la loro valenza a livello discriminante nel forense. Tale decisione si è basata sulle due ipotesi che, all'analisi finale, si sono dimostrate corrette. In primo luogo, l'ipotesi che la più grande quantità di variazione totale la si sarebbe trovata nella regione codificante. La regione codificante ha un tasso evolutivo di circa quattro

volte più basso rispetto alla regione di controllo non codificante (Aquadro e Grennberg, 1983; Horai e Hayasaka, 1990). Tuttavia, la regione codificante è di circa 15 volte più estesa rispetto alla regione di controllo, cosa che la rende una fonte potenzialmente ricca di ulteriori variazioni nella sequenza, per la discriminazione di soggetti che condividono lo stesso D-loop. In secondo luogo, l'ipotesi che i dati di sequenza pubblicati in letteratura potrebbero non essere completamente risolutivi a tale scopo. Al momento sono stati pubblicati diversi genomi completi di DNA mitocondriale, ma un certo numero di essi (Awadalla et al, 1999) è stato dimostrato essere pieno di errori (Macaulay et al, 1999b; Kivisild e Villems, 2000). Gran parte delle conoscenze circa le variazioni all'interno della regione codificante deriva dai numerosi studi mediante gli RFLP, che peraltro coprivano solo un quinto del genoma (Wallace, 1994). Tuttavia molti di questi dati erano di dubbio valore, dal momento che questi siti si collocano tra quelli ad evoluzione lenta, nell'ambito del genoma mitocondriale, a definire "rami globali".

Di recente alcuni ricercatori hanno raccomandato l'aumento della discriminazione forense sequenziando uno o due geni altamente variabili o brevi tratti nella regione del genoma laddove siti variabili multipli erano stati precedentemente identificati (Andreasson et al, 2002; Lee et al, 2002; Lutz-Bonegel et al, 2003). Tale approccio ha però i suoi limiti: innanzitutto, dato il linkage dei polimorfismi dell'mtDNA, anche se un gene venisse osservato come essere altamente variabile, quella variazione non garantisce di risolvere individui che condividono tipi HVI/HVII nella regione di controllo ad evoluzione rapida. Inoltre, focalizzandosi solo su pochi frammenti del

genoma, si potrebbe perdere l'accesso a siti che potrebbero essere sia rari che ampiamente diffusi sul genoma.

2.2 Proposizione ragionata di SNPs utili per la discriminazione forense

Alla luce di queste considerazioni, si focalizzerà l'attenzione sui polimorfismi a singolo nucleotide (SNP), potenzialmente informativi da un punto di vista forense, a patto che: 1) si tratti di sequenze non codificanti site nella regione di controllo, al di fuori delle regioni ipervariabili, e/o di brevi regioni "spaziatrici" attraverso l'intera regione codificante, che non sono codificanti (si estendono, in lunghezza, da una a 30 basi); 2) siano mutazioni silenti (sinonime) nei geni codificanti le proteine; 3) ci sia variabilità all'interno o fra popolazioni di individui; 5) non ci sia ridondanza rispetto ad altri siti selezionati. Relativamente alla neutralità dei siti discriminanti, non tutte le terze posizioni dei codoni dell'mtDNA lo sono. I mitocondri hanno un proprio codice genetico (leggermente differente da quello eucariotico relativo al genoma nucleare), e non vi è, necessariamente, una terza posizione del codone degenerata quattro volte. E' possibile avere una variazione non sinonima nella sequenza aminoacidica, che deriva da transversioni nelle terze posizioni. Al contrario, è anche possibile avere mutazioni in prima posizione ed avere ancora lo stesso aminoacido codificato (vale a dire, ci sono due tRNA che codificano per gli aminoacidi leucina e serina). Ci si può attendere comunque che la maggior parte della variazione non sinonima nel genoma mitocondriale sia anch'essa neutrale (vedere Kimura, 1983). Ad esempio, due dei tre SNPs che sono polimorfismi marcatori del superaplogruppo (vedere oltre) Eurasiatico

N (8701 e 10398) rappresentano variazioni in prima posizione del codone che risultano in un cambiamento da treonina ad alanina. Dati i milioni di individui nel mondo che portano queste mutazioni, ci si può attendere che la tipizzazione di questi SNPs non anticiperebbe alcuna informazione con implicazione clinica. Se uno SNP non sinonimo, o uno SNP relativo ad un tRNA o ad un rRNA è trovato essere di eccezionale utilità forense, si dovrà valutare, in base al singolo caso, se sussistano dei validi motivi per per fare un'eccezione alla regola ferrea della selezione degli SNPs. La valutazione di tali siti potrebbe essere basata sulla loro caratterizzazione in letteratura e sull'ampiezza della loro variazione nella popolazione.

Capitolo 3

MATERIALI E METODI

3.1 Il campione

L'approccio per identificare SNPs utili per risolvere i tipi comuni HVI/HVII si è basatosul sequenziamento dell'intero genoma mitocondriale di vari soggetti Caucasici. Per effettuare la ricerca sono stati prelevati 40 campioni da altrettanti individui provenienti da tutta Italia (fig.3.1).

NORD	CENTRO	SUD	ISOLE
9	9	16	6

Fig. 3.1 Campionamento degli individui analizzati. I colori dello schema corrispondono alle zone, nelle quali è possibile individuare anche le regioni di appartenenza.

Gli individui selezionati, adulti di età compresa tra i 20 ed i 40 anni, appartengono a famiglie residenti in Italia, ed in particolare nelle regioni indicate da almeno 3 generazioni. I campioni sono stati forniti dalla Polizia di Stato; per tutti, indistintamente, è stato garantito l'anonimato per tutte le fasi della procedura.

3.2 L'estrazione del DNA totale

L'estrazione del DNA totale è stata effettuata a partire dai suddetti campioni di sangue (prelevati per venopuntura) con varie tecniche, in relazione alla quantità disponibile. Per i campioni più abbondanti (2.5-5 ml) è stato utilizzato il kit della QIAGEN *Blood & Cell culture DNA Midi*, mentre per quelli di cui erano disponibili solo 100 μl di sangue è stato utilizzato il kit della QIAGEN *Blood & Cell culture DNA Mini*; in entrambi i casi si è fatto riferimento ai rispettivi protocolli consigliati dalla casa produttrice.

3.3 Quantificazione del DNA

Per valutare la qualità del DNA ottenuto, è stata effettuata una lettura spettrofotometrica sui primi campioni studiati, ed una con PCR (vedere oltre) quantitativa sui medesimi campioni, al fine di un controllo interno, nonché sui campioni rimanenti **(**P*CR real time 7500ABI*). I risultati, sebbene più precisi, non si

discostano in maniera significativa da quelli precedentemente osservati allo spettrofotometro.

Sulla base dei risultati ottenuti, a partire dagli stock di DNA, sono state preparate diluizioni in TE (Tris Edta) (10, 0.1, pH8) di 1 ng/µl e 0.1 ng/µl per ogni campione.

3.4 L'amplificazione

Le strategie per amplificare e sequenziare il genoma mitocondriale variano ampiamente tra i diversi laboratori che hanno pubblicato dati relativi all'intera sequenza (Levin et al, 1999; Torroni et al, 2001; Levin etal, 2003). Si è deciso di dividere il DNA mitocondriale in 24 frammenti e di utilizzare come punto di partenza le 24 coppie di oligonucleotidi riportate nell'articolo di Rieder (Rieder et al, 1998; Ingman et al, 2000), a cui ne sono stati aggiunti altri per aiutare a chiarire la zona del D-loop, particolarmente difficile da sequenziare (tab. III.1, figg. 3.2 e 3.3). I frammenti ottenuti si sovrappongono parzialmente alle etremità; sono stati utilizzati sia i primers forward che i relativi reverse, al fine di sequenziare entrambi i filamenti. In alcuni casi, laddove si è resa difficoltosa la copertura di entrambi i filamenti, si è proceduto a sequenziare in doppio lo stesso filamento: ciò al fine di poter comunque fare affidamento su un'informazione di sequenza relativa ad un solo filamento, in

porzioni significative del genoma mitocondriale, fonte potenziale di “mutazioni fantasma” (Bandelt et al, 2002; Yao et al, 2003).

Primer	Sequenza	Estremità 5’	Estremità 3’
1F	5'- CTC CTC AAA GCA ATA CAC TG -3'	592	611
1R	5'- TGC TAA ATC CAC CTT CGA CC -3'	1430	1411
2F	5'- CGA TCA ACC TCA CCA CCT CT -3'	1226	1245
2R	5'- TGG ACA ACC AGC TAT CAC CA -3'	2026	2007
3F	5'- GGA CTA ACC CCT ATA CCT TCT GC -3'	1830	1852
3R	5'- GGC AGG TCA ATT TCA CTG GT -3'	2688	2669
4F	5'- AAA CCT TAC CCC GCC TGT TT -3'	2480	2499
4R	5'- AGG AAT GCC ATT GCG ATT AG -3'	3365	3346
5F	5'- TAC TTC ACA AAG CGC CTT CC -3'	3150	3169
5R	5'- ATG AAG AAT AGG GCG AAG GG -3'	3981	3961
6F	5'- TGG CTC CTT TAA CCT CTC CA -3'	3777	3796
6R	5'- AAG GAT TAT GGA TGC GGT TG -3'	4679	4660
7F	5'- ACT AAT TAA TCC CCT GGC CC -3'	4466	4485
7R	5'- AAT GGG GTG GGT TTT GTA TG -3'	5443	5424
8F	5'- CTA ACC GGC TTT TTG CCC -3'	5238	5255
8R	5'- ACC TAG AAG GTT GCC TGG CT -3'	6050	6031
9F	5'- GAG GCC TAA CCC CTG TCT TT -3'	5835	5854
9R	5'- ATT CCG AAG CCT GGT AGG AT -3'	6661	6642
10F	5'- CTC TTC GTC TGA TCC GTC CT -3'	6450	6469
10R	5'- AGC GAA GGC TTC TCA AAT CA -3'	7334	7315
11F	5'- ACG CCA AAA TCC ATT TCA CT -3'	7129	7148
11R	5'- CGG GAA TTG CAT CTG TTT TT -3'	8114	8095
12F	5'- ACG AGT ACA CCG ACT ACG GC -3'	7909	7927
12R	5'- TGG GTG GTT GGT GTA AAT GA -3'	8816	8797
13F	5'- TTT CCC CCT CTA TTG ATC CC -3'	8602	8621
13R	5'- GTG GCC TTG GTA TGT GCT TT -3'	9416	9397
14F	5'- CCC ACC AAT CAC ATG CCT AT -3'	9211	9230
14R	5'- TGT AGC CGT TGA GTT GTG GT -3'	10149	10130
15F	5'- TCT CCA TCT ATT GAT GAG GGT CT -3'	9967	9989
15R	5'- AAT TAG GCT GTG GGT GGT TG -3'	10858	10839
16F	5'- GCC ATA CTA GTC TTT GCC GC -3'	10653	10672
16R	5'- TTG AGA ATG AGT GTG AGG CG -3'	11511	11492
17F	5'- TCA CTC TCA CTG CCC AAG AA -3'	11295	11314
17R	5'- GGA GAA TGG GGG ATA GGT GT -3'	12095	12076
18F	5'- TAT CAC TCT CCT ACT TAC AG -3'	11929	11948
18R	5'- AGA AGG ATA TAA TTC CTA CG -3'	12793	12774
19F	5'- AAA CAA CCC AGC TCT CCC TAA -3'	12551	12571
19R	5'- TCG ATG ATG TGG TCT TTG GA -3'	13526	13507
20F	5'- ACA TCT GTA CCC ACG CCT TC -3'	13320	13338
20R	5'- AGA GGG GTC AGG GTT GAT TC -3'	14287	14268
21F	5'- GCA TAA TTA AAC TTT ACT TC -3'	14081	14100
21R	5'- AGA ATA TTG AGG CGC CAT TG -3'	15017	14998
22F	5'- TGA AAC TTC GGC TCA CTC CT -3'	14837	14856
22R	5'- AGC TTT GGG TGC TAA TGG TG -3'	15997	15978
23F	5'- TCA TTG GAC AAG TAG CAT CC -3'	15792	15811
23R	5'- GAG TGG TTA ATA GGG TGA TAG -3'	31	11
24F	5'- CAC CAT CCT CCG TGA AAT CA -3'	16401	16420
24R	5'- AGG CTA AGC GTT TTG AGC TG -3'	794	775
26F	5'- AAG CAA GTA CAG CAA TCA ACC -3'	16202	16222
26R	5'- TTT GAT GTG GGT TGG GTT TT -3'	16182	16163
27F	5'- TTC TGG CCA CAG CAC TTA AA -3'	318	337
27R	5'- GGC TGT GCA GAC ATT CAA TT -3'	262	243

Tab. III.1 Elenco dei *primers* scelti per l’analisi. F (forward), R (reverse); oltre ai *primers* del lavoro di Rieder, sono riportati alcuni *primers* aggiuntivi che mappano nella zona del D-loop e sono stati utilizzati su alcuni campioni (vedi testo).

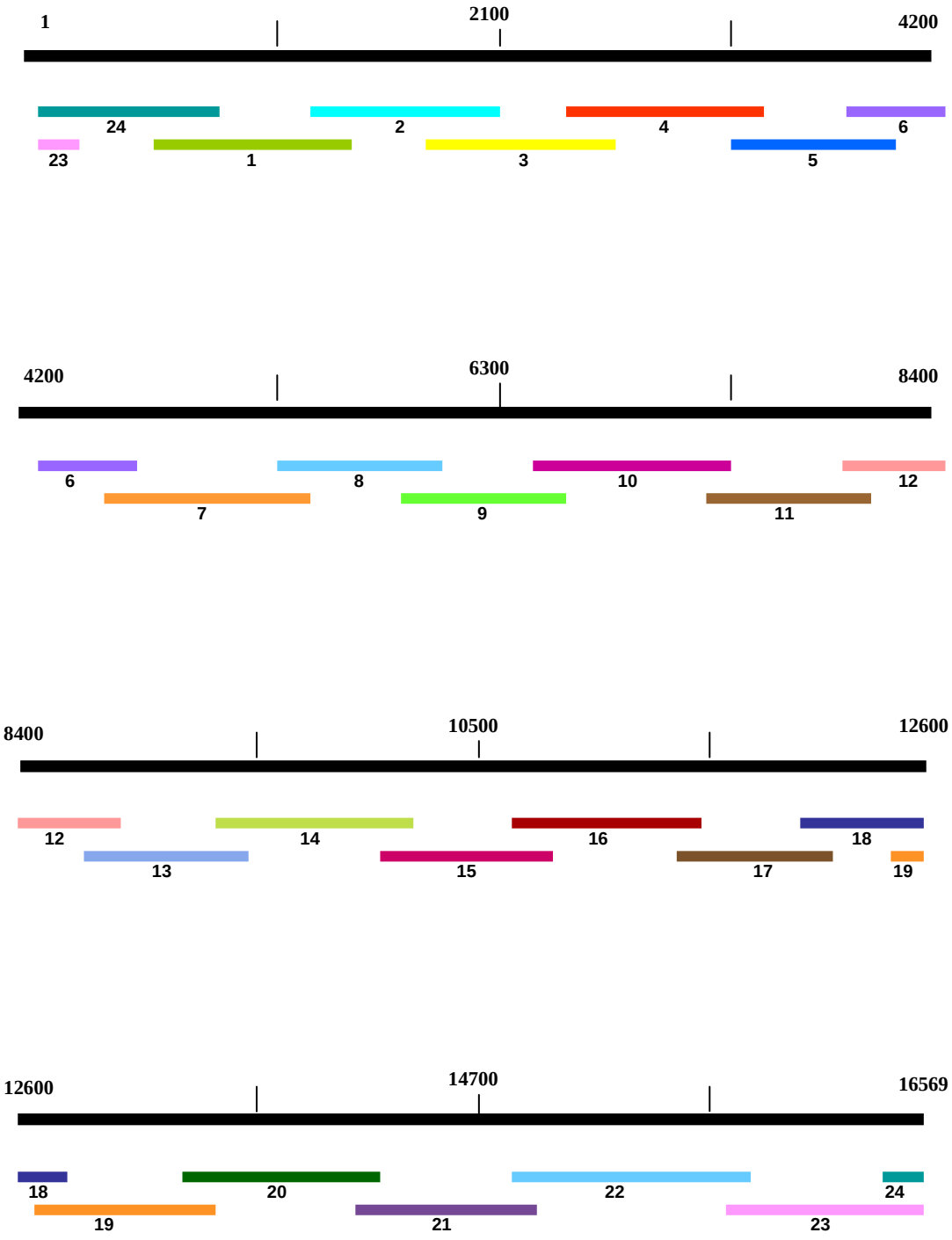


Fig. 3.2 Mappa lineare dei *primers*. Sono indicate le posizioni delle coppie di *primers* sulla mappa del DNA mitocondriale (in nero); è possibile notare che tutte le coppie si sovrappongono per almeno 150 paia di basi.

Come mostrato nella figura 3.2 i frammenti di DNA ottenuti sono studiati in maniera tale da sovrapporsi alle estremità per almeno 150 bp (paia di basi o base

pairs); ciò consente di garantire la copertura totale della molecola dopo l'assemblaggio. Per praticità ci si riferirà, da ora in avanti, ai vari ampliconi utilizzando il numero progressivo dei *primers* corrispondenti.

Tutti i *primers* sono stati verificati allineandoli con la sequenza rivisitata del DNA mitocondriale (Andrews R.M. et al, 1999: sequenza di "Cambridge" rivista o rRCS, numero di accesso iniziale: AC_000021.2; numero di accesso attuale NC_012920.1) e dove necessario sono state apportate le modifiche descritte nella tabella III.2.

Nome	Primer originario	Primer modificato
18R	**5'- AGA AGG TTA TAA TTC CTA CG -3'**	**5'- AGA AGG ATA TAA TTC CTA CG -3'**
20R	**5'- AGA GGG GTC AGG GTT CAT TC -3'**	**5'- AGA GGG GTC AGG GTT GAT TC -3'**
24F	**5'- CAC CAT TCT CCG TGA AAT CA -3'**	**5'- CAC CAT CCT CCG TGA AAT CA -3'**

Tab. III.2 Schema dei *primers* modificati. In rosso sono indicati i nucleotidi sostituiti.

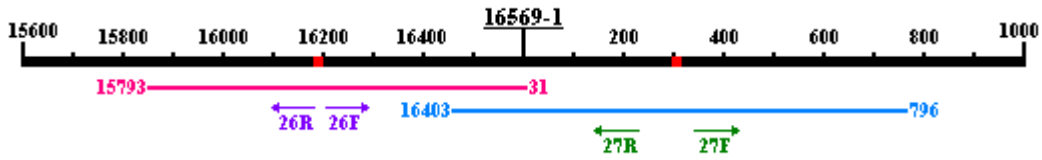


Fig. 3.3 Posizione degli inneschi 26F, 26R, 27F, 27R. La figura illustra nel dettaglio l'utilità dei *primers* supplementari per la zona del D-loop; è possibile osservare che affiancano le regioni in cui sono presenti i tratti di poli-C soggetti ad eteroplasmie di lunghezza (evidenziati in rosso e discussi nel capitolo successivo); l'innesco 27R è più lontano dalla zona interessata, ma corrisponde al migliore ottenibile tra la posizione 200 e la posizione 300.

Per tutti i *primers* è stata calcolata la temperatura di Melting, vale a dire la temperatura di fusione di ciascun oligonucleotide, che dipende dalla sua lunghezza e

dalla composizione in basi secondo la formula: $\mathbf{T_m = 16.6 \log C_s + 41 f_{GC} + 81.5.}$ Il calcolo è stato effettuato sia manualmente che tramite il programma Primer3 e i risultati sono riportati nella tabella III.3.

E' possibile osservare che tutte le coppie di *primers* hanno un T_m simile, che permette di utilizzare una temperatura di annealing di 61°C, tranne le coppie 8FR, 18FR e 21FR. Per queste coppie è stato effettuato un esperimento per la messa a punto della temperatura di annealing tramite PCR a gradiente di temperatura. I parametri di PCR utilizzati sono quelli riportati più avanti, ed il gradiente scelto va da 50°C a 61°C, intervallo che secondo i calcoli doveva comprendere la temperatura ottimale). I risultati sono illustrati in figura 3.4.

Primer	Sequenza	T_m	T_m Primer3
01F	CTCCTCAAAGCAATAACACTG	60,09	53,52
01R	TGCTAAAATCCCACCTTCGACC	60,11	60,07
02F	CGATCAAACCTCACACCACCTCT	60,13	60,11
02R	TGGACAAAACCCAGCTACACCA	60,11	60,11
03F	GGACTAAACCCCTAATACCCTTCTGC	60,12	59,44
03R	GGCAGGTCAATTTACTGGT	60,11	59,97
04F	AAACCTTACCCCGCCTGTTT	60,11	61,92
04R	AGGAATGCCATTGCGATTAAG	60,09	60,06
05F	TACTTCACAAAGCGCCTTCC	60,11	60,39
05R	ATGAAGAATAGGGCGAAGGG	60,11	60,41
06F	TGGCTCCTTTTAACCTCTCCA	60,11	59,81
06R	AAGGAATTATGGATGCGGTTG	60,09	59,78
07F	ACTAATTAACCCCCTGGCCC	60,11	59,16
07R	AATGGGGTTGGTTTTGTATG	60,09	59,41
08F	CTAACCGGCTTTTTGCCC	60,13	60,58
08R	ACCTAGAAGGTTGCCTGGCT	60,13	60,27
09F	GAGGCCTAACCCCCTGTTT	60,13	59,58
09R	ATTCCGAAGCCTGGTAGGAT	60,11	59,92
10F	CTCTTCGTCTGACCGTCCT	60,13	59,40
10R	AGCGAAGGCTCTCAAAATCA	60,09	60,10
11F	ACGCCAAAAATCCATTCACT	60,07	59,43
11R	CGGGAATTGCATCTGTTTT	60,07	59,94
12F	ACGAGTAACACCGAACTACGGC	60,15	60,20
12R	TGGGTGGTTGGTGTAAATGA	60,09	59,67
13F	TTTCCCCCTCTAATTGATCCC	60,11	60,09
13R	GTGGCCTTGGTATGTGCTTT	60,11	60,00
14F	CCCACCCAATCACACATGCCTA	60,11	60,60
14R	TGTAGCCGTTGGAGTTGGTGT	60,11	59,21
15F	TCTCCATCTATTGATGAGGGTCT	60,08	59,07
15R	AATTAGCTGTGTGGGTTGT	60,11	59,85
16F	GCCATACTAGTTTTGCCGC	60,13	59,87
16R	TTGAGAAATGAGTGGAGGCG	60,11	59,98
17F	TCACTCTCACACTGCCCAAGAA	60,11	59,55
17R	GGAGAATGGGGGAATAGGTGT	60,13	60,01
18F	TATCACTCCTCACCTTACGAG	60,07	44,03
18R	AGAAGGAATTACTTCCTACG	60,05	48,03
19F	AAACAACCCAGCTCTCCCTAA	60,10	60,12
19R	TCGATGATGTGGTCTTTGGA	60,09	60,05
20F	ACATCTGTACCCACGCCTTC	60,13	60,00
20R	AGAGGGGTCAGGGTTGATTC	60,13	60,31
21F	GCATAATTAAACTTACCTTC	60,01	44,04
21R	AGAATTTGAGCCCAATTG	60,09	60,06
22F	TGAAACTTCGGCTCAACCTCCT	60,11	59,99
22R	AGCTTTGGGTGCTAATGGTG	60,11	60,13
23F	TCATTGGACAAGTAGCATCC	60,09	56,11
23R	GAGTGGTTAAATGGGTGAATAG	60,08	50,65
24F	CACCATCCCTCGTGAAATCA	60,11	61,88
24R	AGGCTAAGCGTTTTGAGCTG	60,11	59.79
26F	AAGCAAGTACAGCAATCAACC	60,08	57,05
26R	TTTGATGTGGGTTGGGTTT	60,07	60,07
27F	TTCTGGCCACAGCACTTAAA	60,09	59,46
27R	GGCTGTGCAGACATTCAATT	60,09	58,73

Tab. III.3 Tabella delle temperature di Melting. Sono evidenziati i *primers* i cui calcoli presentano incongruenze.

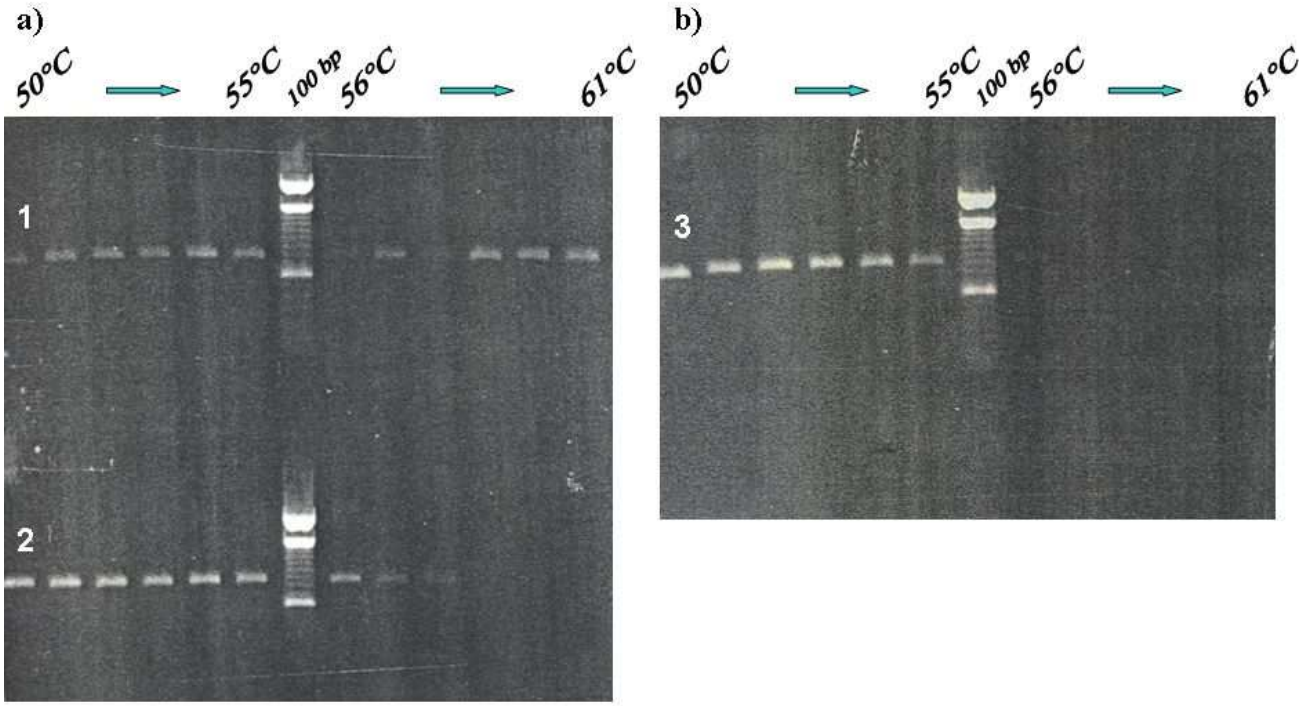


Fig. 3.4 Gel di verifica dei risultati della PCR a gradiente di temperatura. La riga 1 del gel rappresenta i risultati dell'amplicone 8FR, la riga 2 quelli dell'amplicone 18FR e la riga 3 quelli dell'amplicone 21; è possibile osservare che il risultato indica chiaramente una temperatura di amplificazione ottimale di 55°C per gli ampliconi 18 e 21, mentre per l'amplicone 8 il risultato non è definitivo (vedi testo).

In base ai risultati sono state scelte T_m di 59°C per la coppia 8FR e di 54°C per le coppie 18FR e 21FR.

Infine, tutte le coppie sono state verificate anche a livello sperimentale, tramite test su due campioni scelti come riferimento del progetto. Tutti i dati teorici sono stati confermati dalle prove effettuate, con l'eccezione di quelli ottenuti sulla coppia 8FR, nel qual caso i dati sperimentali indicano una variabilità nella resa dell'amplificazione in relazione ai campioni utilizzati (paragrafo 3.4.2).

3.4.1 I parametri

Le amplificazioni sono state effettuate in termociclatori *PCR 9600 e 9700 ABI*. E' stata posta un'attenzione particolare nella scelta del numero di cicli, mantenuto ragionevolmente basso per evitare l'accumulo di errori durante la sintesi.

La miscela di reazione per ogni amplicone conteneva il DNA estratto e quantificato, in ragione di 10 µl (0.1 ng/µl); AmpliTAqGold® DNA polymerase (0.6 µl – 5 U/µl); Buffer II 10X (100 mM Tris – HCl, pH 8.3, 500 mM KCl), 5 µl; $MgCl_2$ (25 mM), 3 µl; dNTPs mix (10 mM), 4 µl; Primer F (10 pmol/µl), 2.5 µl e Primer R (10 pmol/µl), 2.5 µl. Il tutto è stato portato ad un volume finale di 50 µl.

Le condizioni di amplificazione sono state le seguenti (con le eccezioni precedentemente indicate): 10 minuti a 95°C (attivazione della Taq Gold), seguiti da 32 cicli di 94°C per 30 secondi, 61°C per 30 secondi e 72°C per 2 minuti. L'estensione finale è stata eseguita a 72°C per 10 minuti.

3.4.2 Amplificazione della zona 8FR

Come accennato precedentemente, la coppia di *primers* 8FR non funziona in maniera costante nelle stesse condizioni sperimentali, ma varia a seconda dei campioni. In particolare per la maggior parte dei DNA non si è riusciti ad ottenere alcun prodotto di amplificazione; nei rari casi in cui l'amplicone è visibile su gel, la sua quantità non è costante. Inoltre, anche in questi casi, il sequenziamento dell'amplicone 8 è stato particolarmente problematico e le sequenze ottenute non

ottimali; risultati di questo tipo ci hanno portato a formulare due ipotesi. Il tratto di sequenza coinvolto (nucleotidi 5238 – 6050), sul quale già altri studi hanno evidenziato problemi nel sequenziamento, potrebbe formare strutture secondarie con altre zone del DNA mitocondriale o all'interno della zona stessa, risultando in entrambi i casi non accessibile. Per saggiare la prima ipotesi è stata effettuata un'amplificazione con i *primers* 7F e 9R, immediatamente adiacenti e comprendenti la zona di interesse. Nel caso in cui le strutture secondarie si formino all'esterno del tratto analizzato, una volta ottenuti gli ampliconi "tripli", il tratto stesso si potrà sequenziare con i *primers* 8F e 8R. I parametri di PCR sono i medesimi degli altri esperimenti, salvo il tempo di estensione portato da 2 a 4 minuti. Per effettuare questa verifica sono stati scelti otto campioni, quattro dei quali negativi e quattro positivi al test su gel d'agarosio. Di due dei quattro positivi, inoltre, era già stata ottenuta una sequenza discreta, confrontabile con le eventuali nuove. I risultati finora ottenuti sembrano confermare la nostra prima ipotesi, ma gli studi stanno proseguendo su tutti i campioni, al fine di definire più nel dettaglio l'eziologia di quanto indicato.

3.4.3 Purificazione dei prodotti PCR

Per eliminare dagli amplificati il tampone, i dNTPs, la Taq ed i sali, sono state utilizzate due tecniche differenti, entrambe basate su un sistema di membrane a setaccio molecolare.

Per i primi campioni è stato utilizzato il sistema di colonnine *YM 100 MWC* della Millipore, come descritto dal fornitore.

Per i campioni successivi sono state utilizzate le piastre per purificazione della Millipore, *MultiScreen® PCR$_{\mu 96}$*, sulle reazioni di PCR portate ad un volume finale di 100 µl con dH_2O; il DNA è stato fatto legare alla membrana mediante applicazione del vuoto (~20 inHg) per circa 12 minuti, o comunque finché la membrana della piastra risultava asciutta, ma non secca; veniva quindi effettuato un lavaggio con 100 µl di dH_2O applicando il vuoto (~20 inHg) per 10 minuti circa ed i campioni venivano infine eluiti con 40 µl di dH_2O (acqua sterile pretrattata con UV), in un volume finale pari alla metà circa del volume iniziale della reazione. La soluzione recuperata veniva trasferita in piastre *Microtest U – Bottom 3077* (Falcon) e conservata a – 20°C.

3.4.4 La verifica su gel

Al fine di stabilire la purezza e la dimensione dei prodotti purificati, questi ultimi sono stati verificati tramite corsa elettroforetica attraverso gel di agarosio allo 1% (preferibilmente NuSieve/GTG della Seakem nel rapporto 3/1), in tampone TBE (Tris borato EDTA) e Bromuro di Etidio. Come marcatori di peso molecolare sono stati utilizzati insiemi di frammenti (ladders) a concentrazione nota, che differiscono in lunghezzaa di 50 o 100bp. In particolare si è scelto di utilizzare le 100bp ladders della INVITROGEN o della M-Medical. La concentrazione dei prodotti di amplificazione è stata stimata tramite analisi comparata, visiva, delle intensità delle bande degli ampliconi con l'intensità delle bande dei frammenti di lunghezza simile, presenti nello standard di peso molecolare. Un esempio dei risultati ottenuti è

presentato in fig. 3.5. In ciascun pozzetto del gel sono stati caricati 10 µl di amplicone (1/5 della reazione di amplificazione).

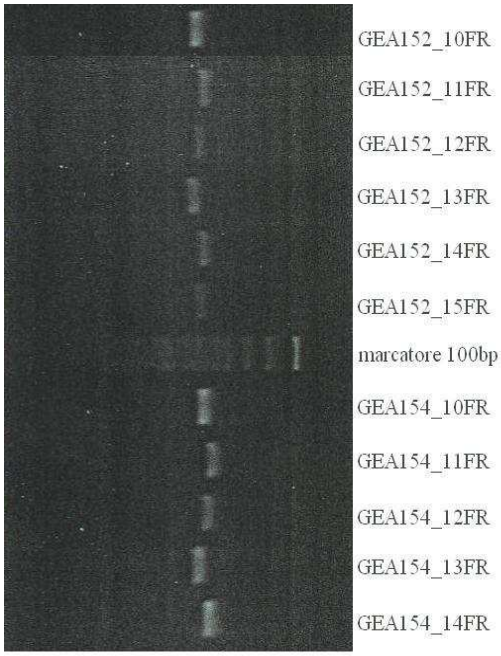


Fig. 3.5 Gel d'agarosio di controllo. La foto illustra un esempio dei risultati del test preliminare dei *primers (campioni di controllo, non utilizzati nel presente studio)*; è possibile notare la diversa altezza delle bande, corrispondente alla differente lunghezza dell'amplicone nei due test utilizzati.

3.5 Il sequenziamento

3.5.1 Le reazioni di sequenza

Per effettuare le reazioni di sequenza è stata scelta la chimica fornita dall'ABI, *BigDyeTerminator versione 3.1*. Per la scelta della quantità di templato si è fatto riferimento alla tabella operativa fornita dall'Applied Biosystems (tab. III.4):

Templato (stampo)		Quantità
Prodotto di PCR	100-200 bp	1-3 ng
	200-500 bp	3-10 ng
	500-1000 bp	5-20 ng
	1000 – 2000 bp	10-40 ng
	>2000 bp	20-50 ng
Singolo filamento		25-50 ng
Doppio filamento		150-300 ng
Cosmidi, BAC		0.5-1.0 µg
Genoma Batterico		2-3 µg

Tab. III.4 Tabella di riferimento per il sequenziamento con la chimica BigDye® Terminator v. 1.1/3.1. Il nostro caso, prodotti di circa 700-900 bp, è evidenziato in giallo.

La miscela di reazione era così composta: Ready Reaction Premix (2.5X), 2 µl; BigDye® sequencing buffer (5X), 3 µl; primer F o R (3.2 pmol/µl), 1 µl; amplicone (~10 ng/µl), 2 µl ed acqua ad un volume finale di 20 µl.

3.5.2 I parametri

I parametri scelti per la PCR della reazione di sequenza dipendono dalla temperatura di Melting dei singoli primers utilizzati. I parametri illustrati di seguito valgono per tutti i *primers* ad eccezione di 8F, 8R, 18F, 18R, 21F, 21R, per i quali la temperatura di annealing è stata adattata a 54°C per i primi due e a 50°C per i successivi.

Denaturazione iniziale di 1 minuto a 96°C, seguita da 25 cicli di 10 secondi a 96°C, 5 secondi a 56°C e 4 secondi a 60°C.

3.5.3 Purificazione delle reazioni di sequenza

Per eliminare dalla reazione ottenuta tutti i reagenti in eccesso, si sono utilizzate delle colonnine riempite di una resina a setaccio molecolare (sephadex – G50 superfine) previamente idratata. Il protocollo utilizzato è lo stesso fornito dalla ditta Applied Biosystems per le colonnine Centrisep, utilizzate in aggiunta alle precedenti, "home made".

Dopo la purificazione, i campioni sono stati essiccati in una speed-vac (SpeedVac Concentrator- SAVANT) e sono stati conservati a -20°C fino al caricamento sul sequenziatore.

3.5.4 L'elettroforesi di sequenziamento

Prima di effettuare l'analisi i campioni sono stati risospesi in 2 μl di dH_2O e denaturati in 23 μl di HiDi Formammide dell'Applied Biosystems.

Il sequenziamento dei campioni è stato effettuato tramite il sequenziatore 3130*xl* ABI situato presso i laboratori della Polizia Scientifica. Per i primi campioni è stato utilizzato il polimero POP-7™ (Applied Biosystems, Foster City, CA) su un capillare di 80 cm, secondo i moduli ed i protocolli di corsa consigliati dalla Ditta. I campioni successivi sono stati analizzati mediante elettroforesi con polimero POP-6™, su un capillare di 50 cm, secondo i moduli ed i protocolli di corsa consigliati e, laddove ritenuto utile, in base alla dimensione degli ampliconi da separare, apportando delle modifiche a livello del tempo di iniezione, di corsa e così via.

3.6 Analisi informatica

La prima analisi dei dati grezzi è stata effettuata tramite il programma fornito dall'Applied Biosystem, *Sequencing Analysis v5.2*.

Le sequenze sono state confrontate con la Sequenza di Riferimento rivisitata di Cambridge (rCRS) (Andrews et al, 1999); quelle relative ai primi campioni sono state analizzate ed assemblate utilizzando l'insieme di programmi denominato Emboss (del Bioinformatic Institute, European Molecular Biology Organization); un secondo controllo, esteso poi ai nuovi campioni, è stato effettuato utilizzando il software Sequencher 4.8 (Gene Codes, Ann Arbor, MI), che, al contrario del precedente, richiede un minore intervento dell'operatore ed è dotato, pertanto, di una maggiore precisione e rapidità. L'assegnazione definitiva, con attenzione ai valori di qualità dei dati ottenuti, è stata effettuata mediante la versione per piattaforma Windows di Phred e Phrap (Codone Code Aligner, Dedham MA, USA). In particolare, Phred perfeziona ulteriormente il *base calling* (cioè l'assegnazione delle basi) ottenuto da Sequencing Analysis e dà una misura della qualità dei dati (da 0 a 50) per ogni base presente nell'elettroferogramma. La qualità della sequenza può essere collegata ad una probabilità di errore, cioè all'accuratezza di ogni assegnazione. Tale relazione è di tipo logaritmico, vale a dire uno score di qualità, dato da Phred, pari a 10, corrisponde ad un'accuratezza nell'assegnazione della base pari al 90%, mentre uno score di 50 corrisponde ad un'accuratezza del 99.99%. Tre criteri vengono utilizzati per generare le misure di qualità in Phred: essi comprendono lo spacing (la distanza)

tra un picco e l'altro, le dimensioni relative dei picchi assegnati e non, e la variazione del segnale tra i picchi assegnati. L'assemblaggio delle sequenze e la generazione della sequenza consenso globale, relativa ad ogni individuo, è stata eseguita mediante Phrap. Phrap utilizza le chiamate delle basi e l'informazione relativa alla qualità, ottenuta da Phred, ed allinea le varie sequenze, che possiedono delle zone di sovrapposizione. Phrap utilizza la qualità della sequenza per generare un consenso dalle chiamate delle basi di più alta qualità nell'assemblaggio finale e genera una qualità ridefinita (range 0-90), in ogni posizione, prendendo in considerazione la qualità del contesto della sequenza e la conferma a mezzo del filamento opposto (Ewing e Green, 1998; Ewing et al, 1998).

L'allineamento multiplo, relativo a particolari tratti di sequenze appartenenti ad individui differenti, è stato effettuato con il programma ClustalX (Thompson et al, 1994).

Per ogni soggetto studiato è stata compilata una lista di polimorfismi, riferiti alla rRCS e, nel caso di polimorfismi facenti parte di sequenze codificanti, il relativo impatto della mutazione a tale livello, cioè se questa fosse sinonima o meno. Come supporto informatico per tale tipo di approccio ci si è avvalsi del software online "MitoAnalyzer" (Lee e Levin, 2002). Si è voluto accertare, per ciascuno dei polimorfismi, se il suo riscontro fosse già stato documentato in letteratura, o se ci si fosse trovati di fronte ad uno nuovo, se le variazioni così individuate fossero associate a condizioni patologiche, e, in caso affermativo, in quale misura. Sulla base

di tutte le informazioni così ricavate si è inoltre verificato, a livello di altri studi, se i polimorfismi in questione fossero già stati individuati come SNPs utili per saggi forensi e/o se potessero concorrere alla messa a punto di tali pannelli.

Capitolo 4

RISULTATI E DISCUSSIONE

4.1 Analisi delle sequenze

In questo studio dei polimorfismi del DNA mitocondriale nella popolazione italiana, sono stati sequenziati 40 campioni provenienti da tutta Italia. Nelle figure che seguono sono mostrati alcuni elettroferogrammi ottenuti, a dimostrazione della qualità dei dati discussi in seguito (figg. 4.1, 4.2, 4.3).

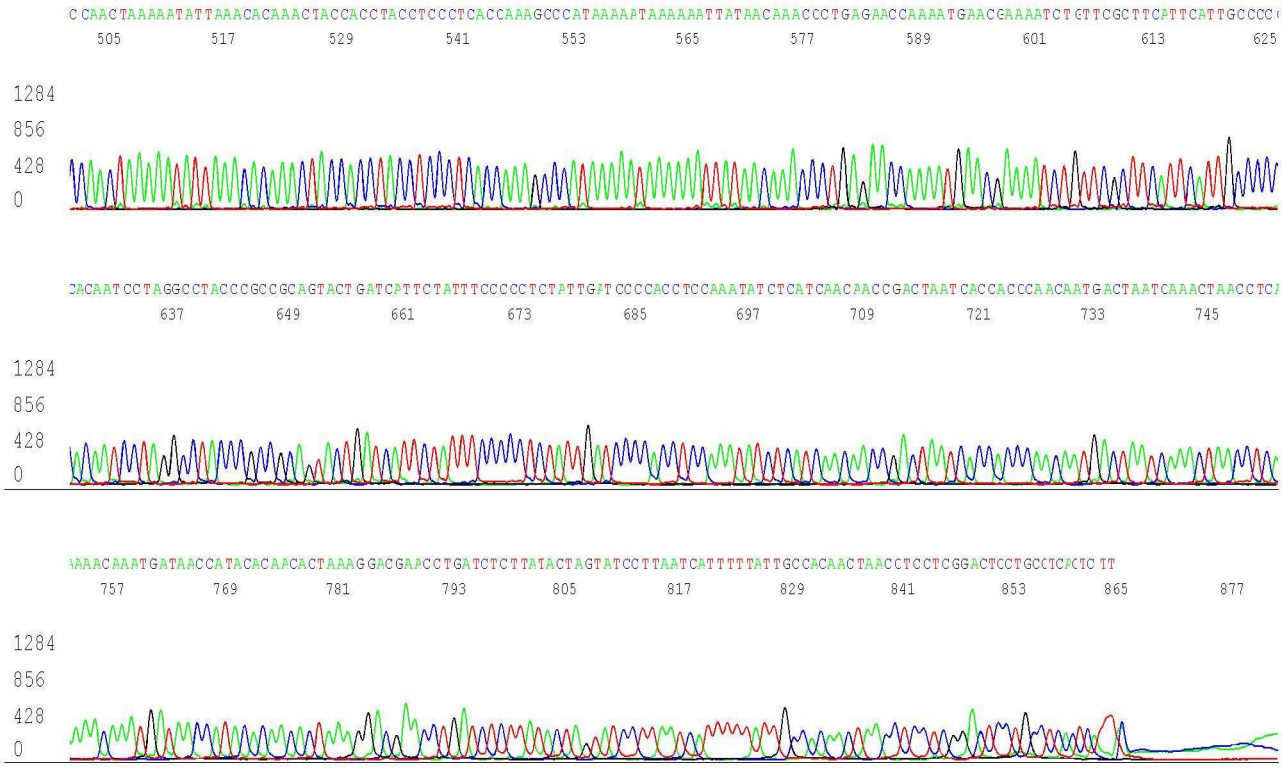


Fig. 4.1 Esempio di elettroferogramma considerato ottimale. Campione test: tratto della sequenza ottenuta con l'innesco 12F.

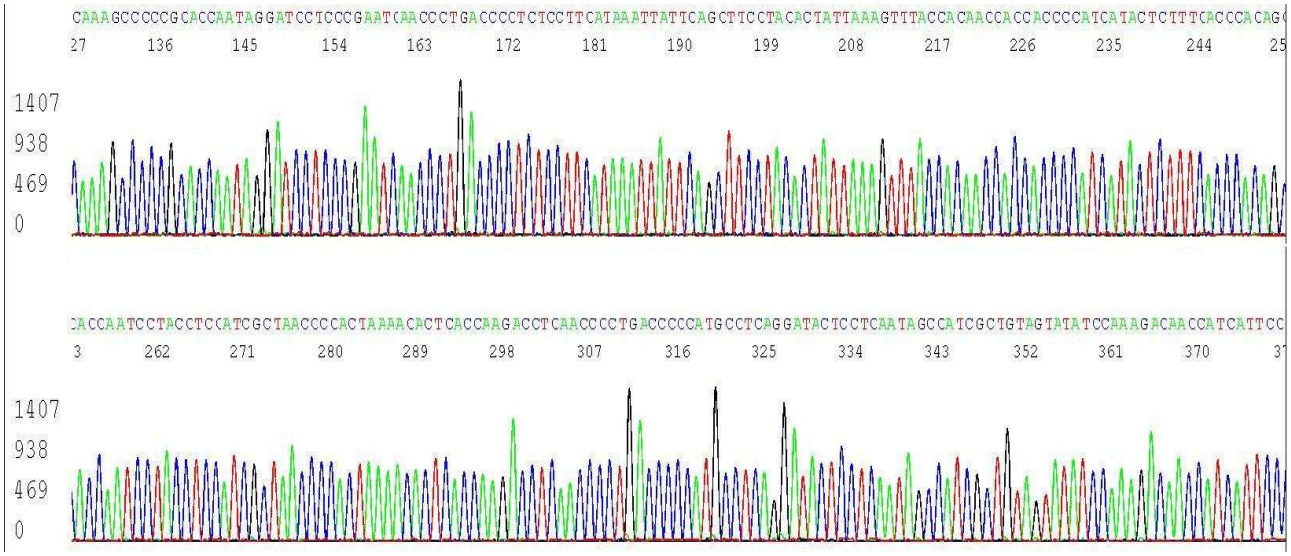


Fig. 4.2 Esempio di elettroferogramma considerato ottimale. Campione test: tratto della sequenza ottenuta con l'innesco 21F.

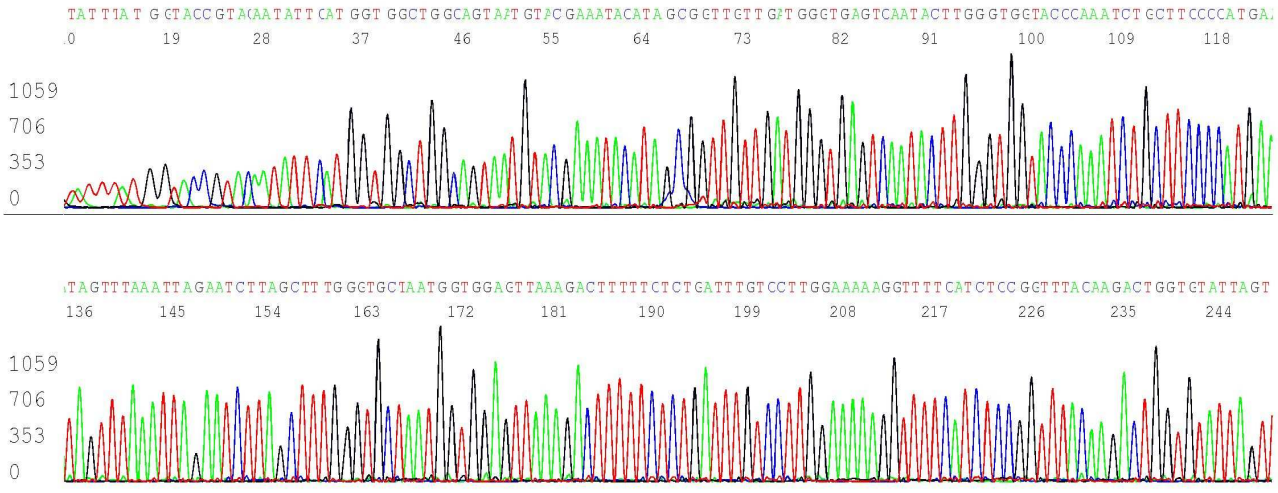


Fig. 4.3 Esempio di elettroferogramma considerato ottimale. Campione GEA59, tratto della sequenza ottenuta con l'innesco 26R.

Per l'analisi delle sequenze, il valore minimo medio accettabile da Phrap per la qualità della sequenza (QV = Quality Value) su tutto il genoma è stato posto pari a

20. In generale, valori di qualità superiori a 20 (che corrispondono ad una massima probabilità di errore dello 1% nell'assegnazione delle basi) possono essere considerati alti per l'interpretazione della sequenza dominante (vedere anche Parson et al, 2004). In molte regioni del genoma mitocondriale si sono ottenuti valori di qualità molto più alti (>50) (fig. 4.4). La variabilità nei valori di qualità dati da Phrap, attraverso il genoma mitocondriale, dipende dalla proporzione di sequenze di conferma e, tipicamente, la qualità più alta si registra vicino alla metà di ogni frammento, nel punto in cui le reazioni di sequenza forward e reverse si sovrappongono. Anche se le sovrapposizioni sono localizzate all'inizio ed alla fine dei frammenti adiacenti, queste regioni possono presentare qualità di sequenza più bassa dovuta alla ridotta intensità del segnale dei filamenti che si sovrappongono ed alla presenza dei "picchi di primer". Altre zone che possono presentare bassi QV sono quelle relative ai cosiddetti "spikes", legati a non ottimale lettura della sequenza in determinati punti (fig. 4.5). L'informazione che complessivamente se ne ricava descrive oggettivamente l'affidabilità dell'assegnazione di ciascuna base, e può fungere come ulteriore supporto nel trattamento di dati di alta qualità.

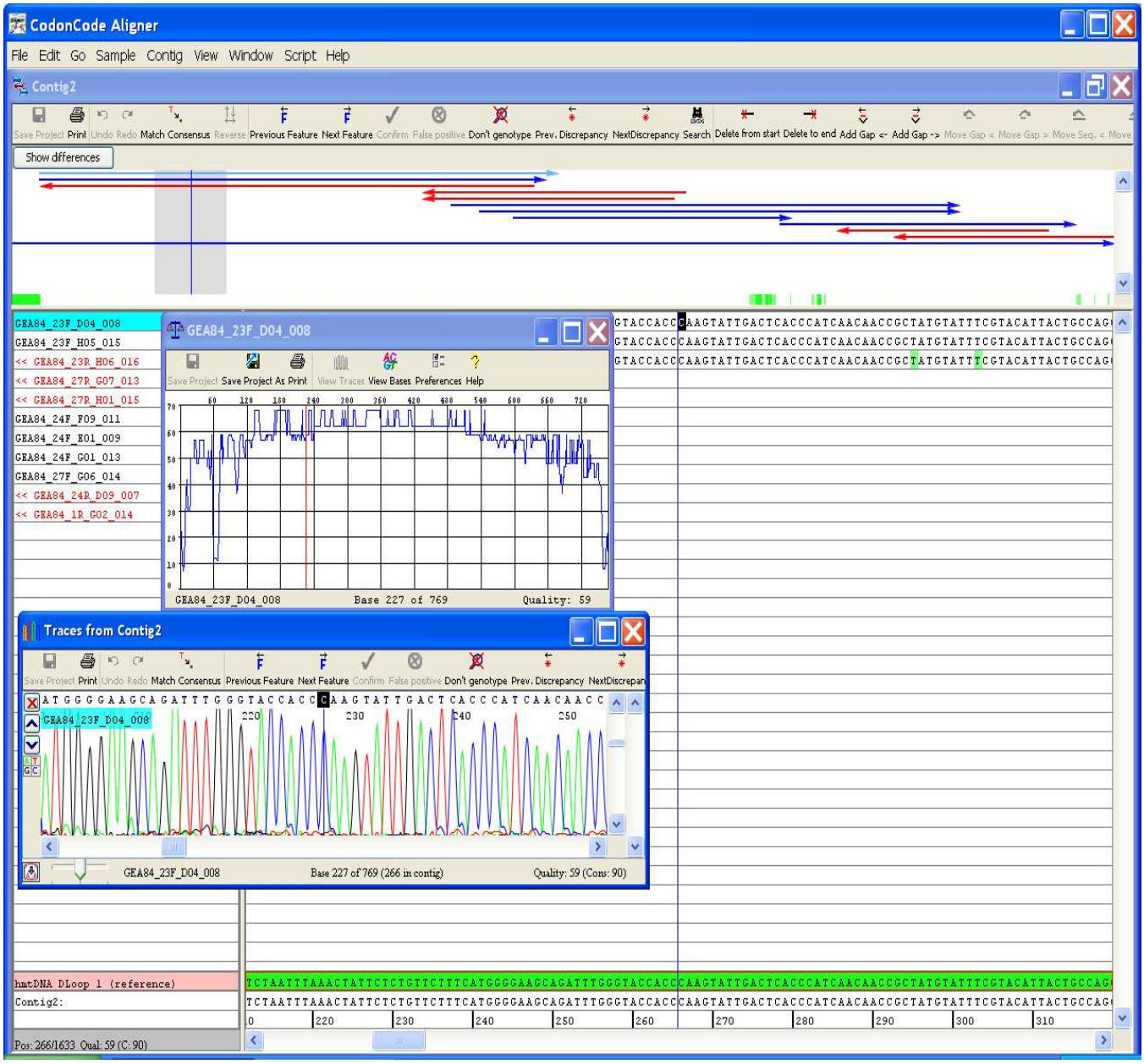


Fig. 4.4 Esempio di alto QV (Quality Value) in un tratto di sequenza.

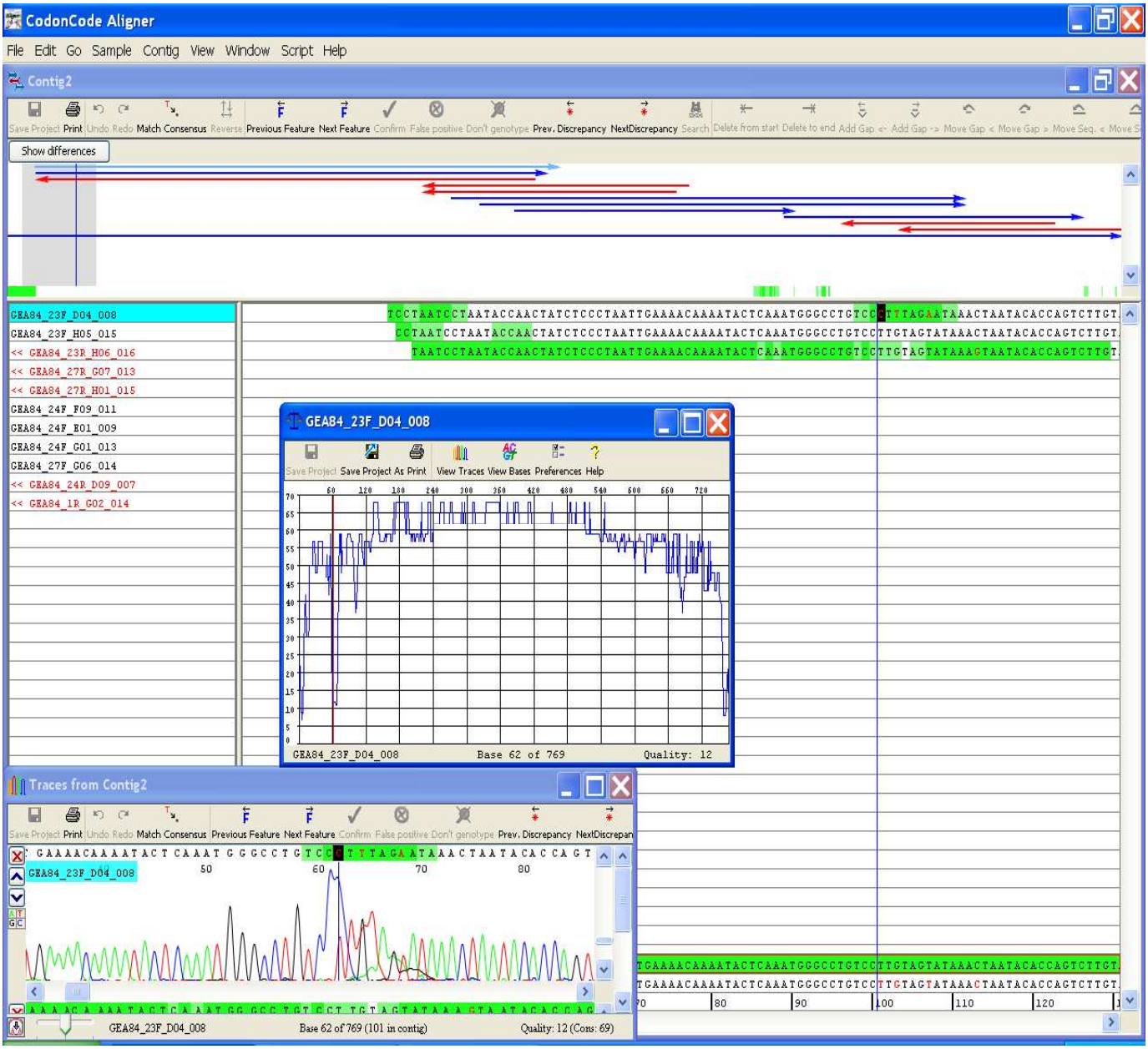


Fig. 4.5 Esempio di basso QV (Quality Value) in un tratto di sequenza.

La sequenza cui si fa riferimento per l'analisi è la sequenza di Anderson aggiornata (numero di accesso GenBank NC_012920.1, dall'08 luglio 2009). Tutti i dati ottenuti sono stati confrontati con gli elenchi presenti sul database **MITOMAP** (www.mitomap.org) aggiornati alla data del 16 maggio 2010 e con l'elenco dei polimorfismi del database **Human Mitochondrial Genome Database** (www.genpat.uu.se/mtDB).

4.2 Eteroplasmie di lunghezza

L'eteroplasmia di lunghezza è stata osservata fin dai primi studi di sequenziamento. L'origine dei polimorfismi di lunghezza sembra imputabile all'instabilità dei tratti poli-C. Infatti normalmente questi tratti sono interrotti da un residuo di timina, guanina o adenina, mentre negli individui che presentano il polimorfismo spesso non c'è interruzione. Il tratto di citosine consecutive ha maggiori probabilità di essere soggetto a "slittamento" durante la replicazione generando così una mistura di sequenze di varia lunghezza. Secondo i dati ottenuti ad oggi è possibile ipotizzare una specificità nell'inserzione di tratti poli-C, poiché le suddette inserzioni sono state osservate solo nelle posizioni 310, 568, 961, 5895, 8281, 9537 e 16189 (Janssen et al, 2006).

L'analisi dei 40 campioni ha rivelato la presenza di polimorfismi associati ad eteroplasmie di lunghezza su quattro segmenti distinti. I dati al riguardo sono stati confermati da sequenze in entrambe le direzioni.

4.2.1 Polimorfismi noti in letteratura

4.2.1.1 Polimorfismo poliC-303

Il polimorfismo di lunghezza più comune (come dimostra il lavoro ed i dati presenti in letteratura) riguarda il tratto poli-C che inizia all'interno della regione HVII in posizione 303 (il polimorfismo è indicato di seguito come poliC-303) (Stewart et al, 2001; Lutz-Bonengel et al, 2004). In riferimento alla sequenza rivisitata di Cambridge, tutti i campioni presentano il polimorfismo, con tipologie variabili, illustrate nella figura e nella tabella sottostanti. La percentuale dell'eteroplasmia di lunghezza varia in ogni campione, come indica la qualità dell'elettroferogramma dopo il segmento poliC. In particolare, il polimorfismo è molto accentuato nei campioni che non presentano la timina in posizione 310 o in cui è presente in condizione di eteroplasmia netta con una citosina (non di rado seguita da eteroplasmia simmetrica C/T). La variabilità osservata nel tratto poliC-303 è mostrata nella figura 4.6, ed è riassunta nella tabella IV.1.

Per ottenere la sequenza completa in ambedue le direzioni sono stati utilizzati come inneschi delle reazioni di determinazione delle sequenze i due oligonucleotidi 27F e 27R, appositamente disegnati (Materiali e Metodi, fig. 3.3; vedere anche fig. 4.7).

Fig. 4.6 Allineamento di 17 sequenze del DNA mitocondriale nel tratto adiacente al poliC-303. La figura mostra la variabilità trovata nel tratto poliC-303.

Tipo	Campioni	Sequenza	Tot.
I	21,28,44,57,59,60,61,62,65,66,70,73,76,80,83,85,96,98,104,105,109,111	CCCCCCC**C**TCCCCC**C***	22
II	58,63,64,67,68,69,71,72,79,81,87,107,108	CCCCCCCTCCCCC**C**	13
III	82,88,100	CCCCCCC**CC**TCCCCC**C**	3
IV	84	$C_{(n)}$ con n ≥ 14	1
V	86	CCCCCCCT**TC**CCCCC	1

Tab. IV.1 Schema della variabilità osservata nel tratto poliC-303. La tabella riassume le combinazioni trovate, specificando il totale delle osservazioni (Tot.) e i nomi dei campioni con l'omissione del codice GEA.

* Tratto poliC-303 presente nella precedente sequenza di riferimento (Accession numer in Gene Bank: NC_001807.4)

I campioni di "Tipo I" sono generati dalla contemporanea inserzione della base C in posizione 309 (309.1C) e 315 (315.1C).

I campioni di "Tipo II" sono generati dalla sola inserzione della C nella posizione 315 (315.1C).

I campioni di "Tipo III" sono generati dalla contemporanea inserzione di due basi C in posizione 309 (309.2C) e di una sola C in posizione 315 (315.1C).

Il campione di “Tipo IV” è definito dalla transizione T→C nella posizione 310. Questa tipologia è molto rara, essendo osservata in 3 su 120 Europei (Lutz-Bonengel et al, 2004).

Il campione di “Tipo V” è definito dalla duplice inserzione di una T e di una C nella posizione 310 (310.2TC), interpretabile, alternativamente, come una inserzione di T in 310 (310.1T) e di C in 315 (315.1C).

Fig. 4.7 Esempio di perdita di risoluzione in HVII ed utilizzo del primer interno 27F per risolvere la sequenza dominante.

4.2.1.2 Polimorfismo poliC-16184

Il polimorfismo di lunghezza poliC-16184 mappa all'interno della regione HVI nel tratto di poliC che inizia al nucleotide 16184 (Bendall et Sykes, 1995; Asari et al, 2007). Otto campioni su 40 (20%) sono interessati dal polimorfismo (fig. 4.8). In diversi campioni in cui è stato individuato il polimorfismo poliC-16184 è presente un'eteroplasmia T/C in posizione 16189. Quindi nello stesso segmento si osserva un'eteroplasmia sia di sequenza che di lunghezza. Anche in questo caso sono stati utilizzati due oligonucleotidi ulteriori, 26F e 26R, per completare il sequenziamento su entrambi i filamenti (Materiali e Metodi, Fig.2.6). Sono state osservate diverse varianti (fig. 4.9), riassunte nella tabella IV.2.

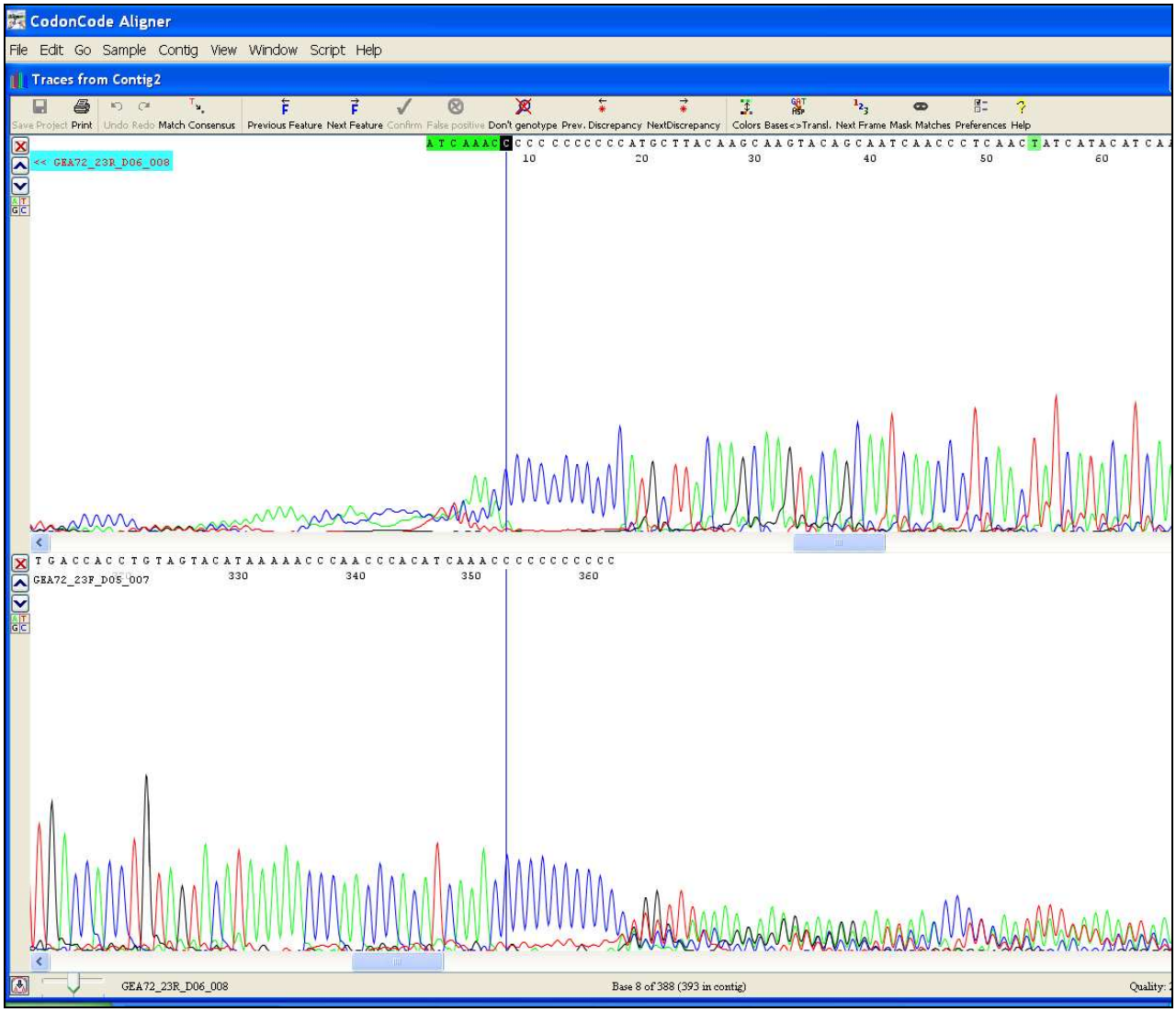


Fig. 4.8 Esempio di perdita di risoluzione in HVI (Primers 23FR).

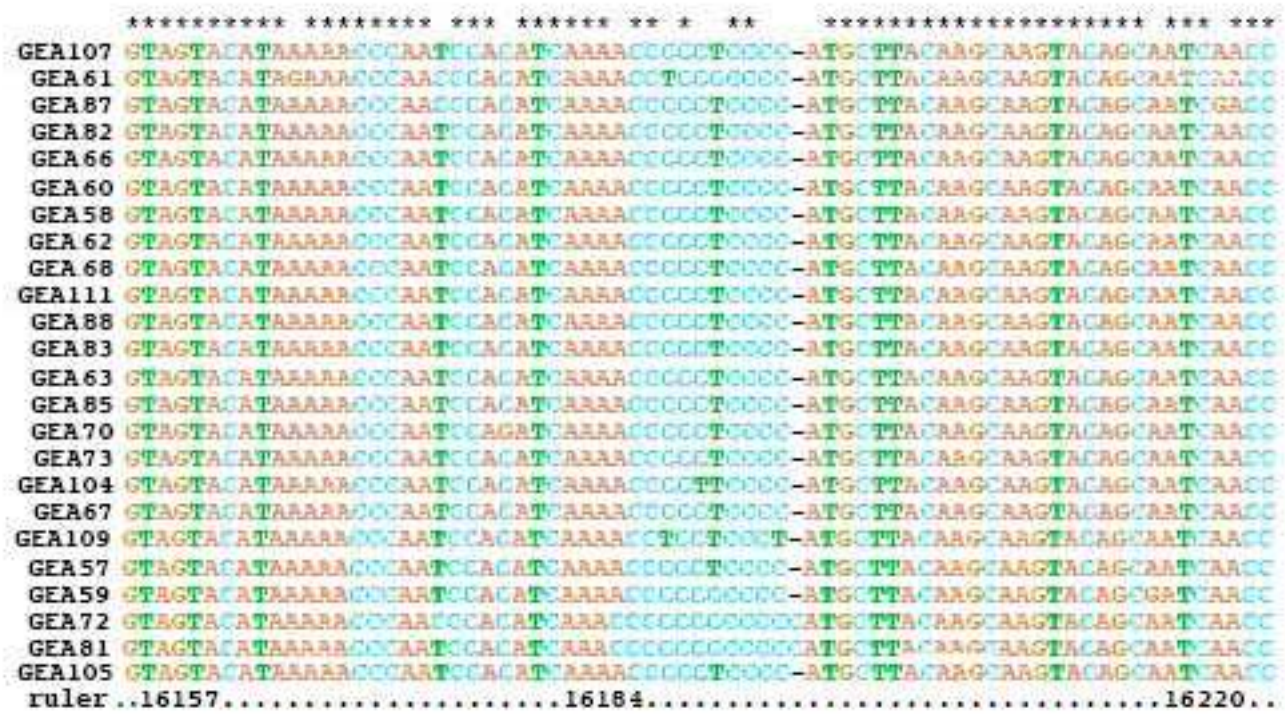


Fig. 4.9 Allineamento di 24 sequenze: la figura mostra il tratto di sequenza contenente il poliC-16184.

Tipo	Campioni	Sequenza	Tot.
I	21,28,44,57,58,60,62,63,64,65,66,67,68,70,71,73,76,80,82,83,84,85,86,87,88,96,98,100, 105,107,108,111	CCCCCTCCCC*	32
II	61,69	CC**T**CC**C**CCCC	2
III	104	CCCC**T**TCCCC	1
IV	79	CCCCCTCC**T**C	1
V	109	CC**T**CCTCCC**T**	1
VI	59,72,81	$C_{(n)}$ con n ≥ 12	3

Tab. IV.2 Schema della variabilità osservata nel tratto poliC-16184. La Tabella riassume le combinazioni trovate, specificando il totale delle osservazioni (Tot.) e i nomi dei campioni con l'omissione del codice GEA.

*Coincidente con l'attuale Sequenza di riferimento (Accession numer in Gene Bank: NC_012920.1)

I campioni di "Tipo I" coincidono con l'attuale sequenza di riferimento, come indicato nella nota.

I campioni di "Tipo II" derivano dalla contemporanea transizione C16186T e T16189C.

Il campione di “Tipo III” si genera in seguito alla transizione C16188T.

Il campione di “Tipo IV” si genera in seguito alla transizione C16192T.

Il campione di “Tipo V” deriva dalla contemporanea transizione C16186T e C16193T.

I campioni di “Tipo VI” si generano, nello specifico: 1)GEA59: in seguito alla transizione T16189C; 2) GEA72: in seguito alla duplice transizione T16189C e A16183C; vi è inoltre l’inserzione di una C in posizione 16193; 3) GEA81, in seguito alla transizione T16189C ed A16183C.

Sequenza	Variante	% europea	% campioni
CCCCCTCCCC	Ref.NC	77.18%	80%
CTCCCCCCCC	A	0.22%	0
CCTCCCCCCC	B	2.46%	5%
CCCTCCCCCC	C	0.67%	0
CCCCCCCCTC	D	0.67%	0
	Varianti rare	7.84%	7.5%
CCCCCCCCCC	8C	10.96%	7.5%
	9C		
	10C		
	11C		
	12C		

Tab. IV.3 Confronto delle frequenze delle varianti osservate nel tratto di poliC-16184 nei DNA mitocondriali europei e nei campioni GEA

Nella tabella IV.3 sono riportate le frequenze delle variazioni polimorfiche del tratto poliC-16184 nei 40 campioni analizzati, confrontata con quella di 447 europei (Chinnery et al, 2005). I risultati del paragone indicano che i dati sono simili. L’unica discrepanza riguarda la variante B che nel nostro campione è sovrarappresentata.

Sono stati effettuati numerosi studi sulla possibile associazione del polimorfismo poliC-16184 con varie patologie umane fra cui il diabete di tipo 2, tuttavia finora non è stato ottenuto nessun dato che dimostri questa correlazione (Chinnery et al, 2005).

4.2.1.3 Polimorfismo poliC-568

Il polimorfismo di lunghezza poliC-568 è stato osservato nella sequenza di DNA mitocondriale GEA62, nel tratto di poliC che inizia in posizione 568 (fig. 4.10). Nella sequenza è presente un'inserzione di almeno cinque citosine (vedi paragrafo successivo): questo numero è variabile poiché la sequenza è polimorfica. E' stata rilevata, inoltre, la presenza della transversione A→C in posizione 574, che pure concorre alla generazione di questo poliC. Un altro modo di descrivere questa variazione, in concomitanza con la precedente, è 573.1C, indi la A (*wild*). Ciò è legato, come appena detto, al particolare assetto della sequenza. Quest'ultima descrizione consente di collocare il campione in uno specifico sottoaplogruppo (vedere oltre). Il problema della nomenclatura delle varianti di lunghezza (cfr anche il Tipo V del poliC-303, visto sopra) è stato affrontato in varie fasi dalla comunità scientifica ed è stato di recente ripreso, alla luce dei canoni di qualità per l'analisi del DNA mitocondriale, stabiliti dall'EMPOP (Bandelt e Parson, 2008): vedere discussione.

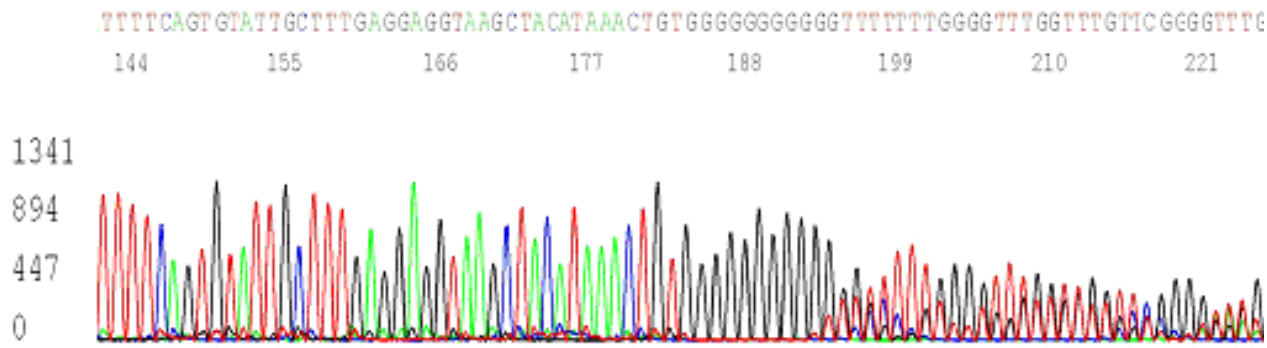


Fig. 4.10 Elettroferogramma della sequenza dell'amplicone 24FR del DNA GEA62. La figura mostra la sequenza nucleotidica polimorfica poliC-568, ottenuta con l'innesco 24R. E' possibile osservare la perdita di risoluzione subito dopo il poliG.

4.2.2 Polimorfismi nuovi

Un polimorfismo di lunghezza non ancora descritto in letteratura è stato individuato nel tratto di poli-C che inizia in posizione 456 all'interno della regione HVII. Un solo campione, GEA68, presenta questa variante (figg. 4.11 e 4.12). All'interno della medesima, è possibile rilevare i seguenti polimorfismi: T455C, T460C e 464A-, che determinano la molteplicità di sequenze osservate. Il DNA mitocondriale di questo individuo presenta, inoltre, il polimorfismo poliC-303, descritto nel paragrafo precedente.

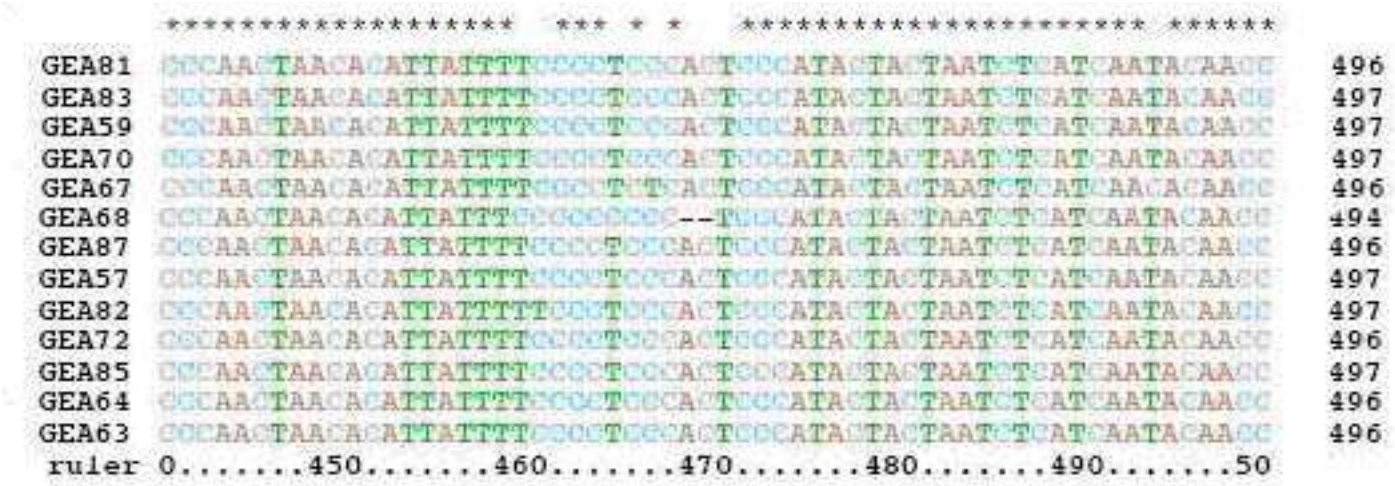


Fig. 4.11 Tratto di sequenza adiacente al polimorfismo poliC-456: la figura mostra l'allineamento di 21 sequenze rappresentative e sintetizza la variabilità trovata.

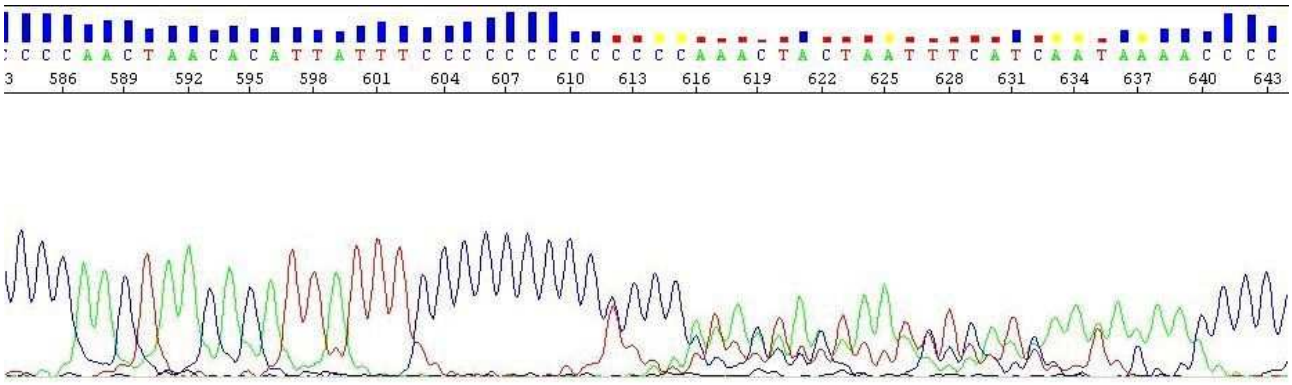


Fig. 4.12 Elettroferogramma del campione GEA68, innesco 24F: dettaglio della zona polimorfica.

4.3 Inserzioni e delezioni

Nei 40 campioni sono state osservate inserzioni e delezioni di singoli nucleotidi oppure tratti di sequenza, come riassunto nella tabella IV.4.

L’inserzione dell’adenina alla posizione 5752 del mtDNA GEA65 è stata riportata in uno studio su una popolazione anziana, con rischio di Alzheimer, ma l’associazione con questa patologia non è ancora stata chiarita (Mostafaie et al, 2004). Essa mappa nella regione non codificante della sequenza in cui è presente l’origine di replicazione O_L (Anderson et al, 1981). Non conoscendo il quadro clinico del soggetto non è possibile ipotizzare un’eventuale associazione di questa inserzione con patologie.

Campione	Inserzione	Delezione
GEA62	568.5CCCCC*	-
GEA63	-	522CA-d
GEA64	-	522CA-d
GEA65	5752.1A	-
GEA68	-	464Ad*
GEA72	16193.1C*	522CA-d
GEA73	14191.1A	
GEA82	-	522CA-d
GEA84	525-CACA	-
GEA85	-	522CA-d

Tab. IV.4 Schema riassuntivo delle delezioni e inserzioni osservate: la posizione riportata, ad eccezione dell’inserzione del campione GEA65, serve solo a dare un’indicazione riguardo alla zona in questione, poiché in tutti i casi si tratta di tratti di sequenze ripetute.

*** vedere paragrafi precedenti.** Per comodità si è omesso di inserire le inserzioni in 309 e 315, di cui si è già discusso.

Di nuova descrizione risulta l’inserzione di una A nella posizione 14191, individuata in GEA73, nonché la delezione di una A nella posizione 464, individuata in GEA68.

Nel tratto di DNA che inizia alla posizione 514 sono presenti ripetizioni CA in tandem. Sono state osservate inserzioni (GEA84) e delezioni (GEA63, GEA64, GEA72, GEA82, GEA85) come già descritto in letteratura.

L'inserzione del tratto di 5C osservato nel mtDNA GEA62 (fig. 4.13) è stata associata, in concomitanza ad altre varianti, tra cui la delezione di una delle ripetizioni CCCCCTCTA, discussa in seguito, ad una patologia multisistemica caratterizzata da diabete giovanile e sordità progressiva, detta MIDD (Maternally Inherited Diabetes and Deafness). Questa inserzione è molto rara nei caucasici, mentre è stata riscontrata più frequentemente in popolazioni geneticamente isolate (Janssen et al, 2006). Da ricordare, infine, le inserzioni di C riscontrate nelle posizioni 309 e 315, da cui originano i polimorfismi di lunghezza sopra discussi. Parimenti, è da ricordare la delezione N3107del, comune a tutti i campioni, così indicata a seguito della nomenclatura della nuova sequenza di riferimento (vedere www.mitomap.org).

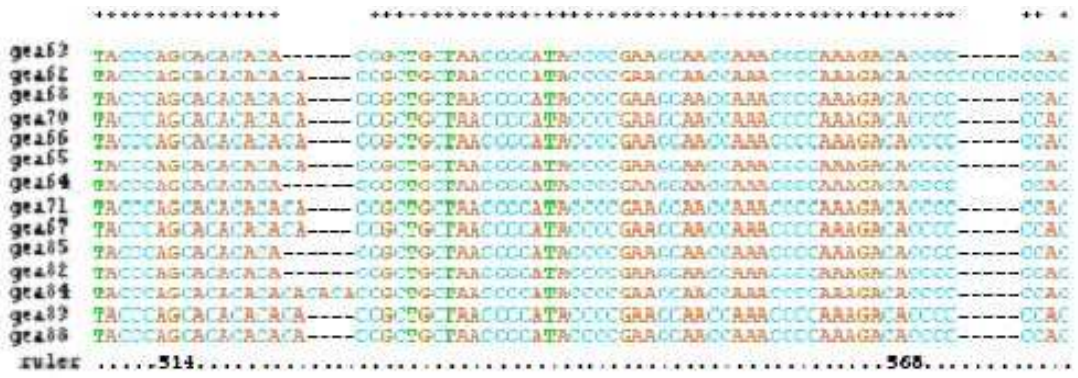


Fig. 4.13 Allineamento di 14 campioni esplicativi delle tipologie di inserzioni e delezioni riscontrate: è possibile osservare il tratto di ripetizioni in tandem CA e le varianti presenti nei campioni, inserzioni e delezioni, nonché l'inserzione di 5C alla posizione 568.

4.4 Eteroplasmie

Nel corso dello studio delle variazioni polimorfiche sono stati osservati due casi di eteroplasmia netta, in cui i segnali negli elettrofrogrammi risultano quasi totalmente sovrapposti (Tab. IV.5).

Eteroplasmia	Gene coinvolto o regione	Campione
6800 A/G	COI	GEA82
16037 A/G	D-loop (HVI)	GEA104

Tab. IV.5 Eteroplasmie evidenti tra quelle individuate nei 40 campioni in esame.

L'eteroplasmia 6800 A/G non è mai stata riportata in letteratura ma è stata descritta la transizione A → G come unica osservazione in Gene Bank (Malyarchuk et al, 2008a), che verrà discussa nel prossimo paragrafo (fig. 4.14).

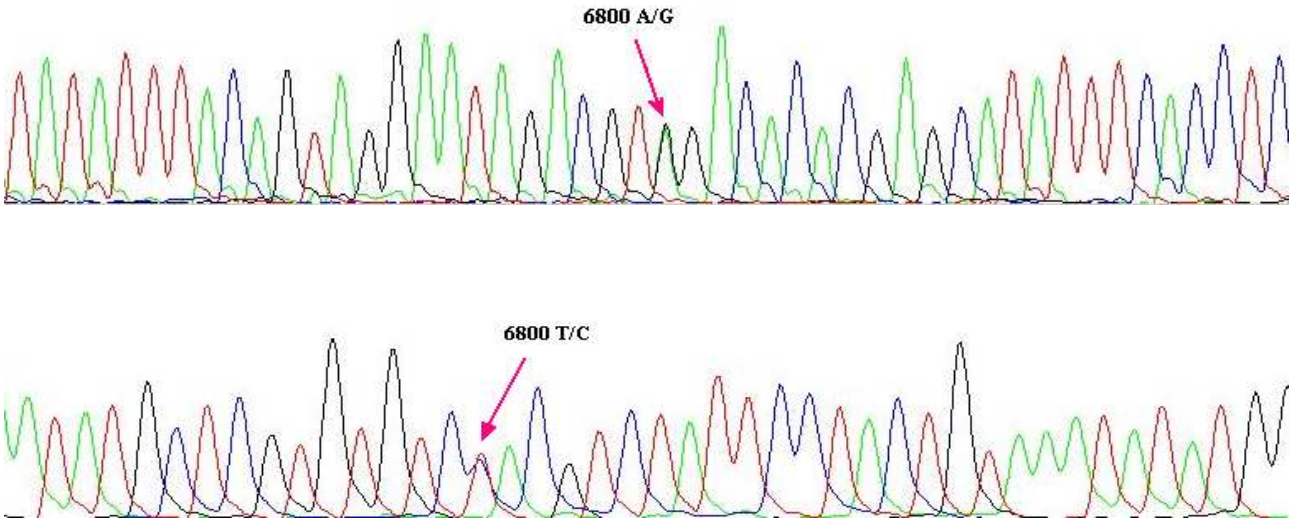


Fig. 4.14 Elettroferogrammi parziali delle sequenze del campione GEA82 innescate con gli oligonucleotidi 10F e 10R. La figura mostra l'eteroplasmia in posizione 6800 visibile su entrambi i filamenti (freccia rosa).

Anche l'eteroplasmia 16037 A/G non è stata indicata in letteratura come tale (Fig. 4.15), ma solo come transizione A→G.

Nella figura 4.15 è evidenziato anche un secondo caso di eteroplasmia in posizione 16054. Durante l'analisi degli elettroferogrammi ottenuti, sono state osservate numerose eteroplasmie simili a questa. Si è deciso di non prenderle in considerazione poiché presentano un rapporto fra le intensità dei due segnali inferiore a 0.25.

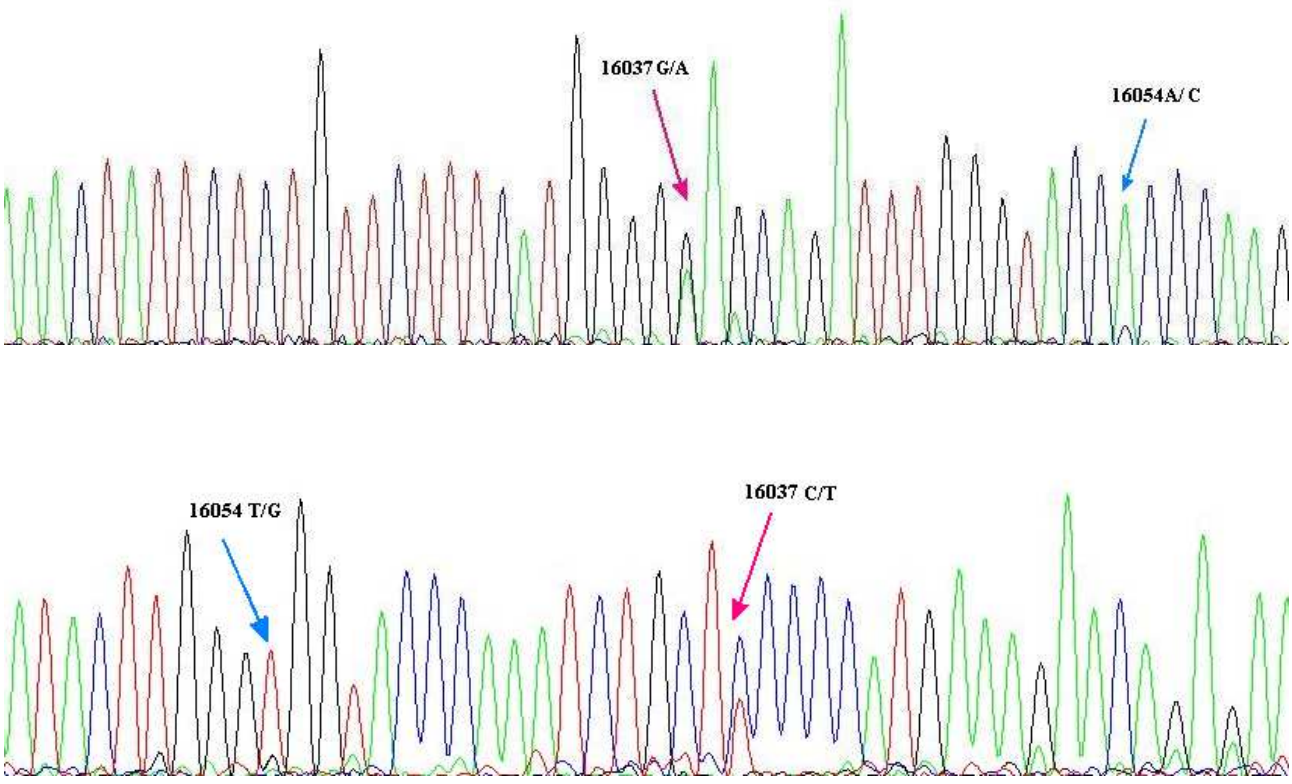


Fig. 4.15 Parte dell'elettroferogramma F e R dell'amplicone 23FR, campione GEA104. La figura mostra l'eteroplasmia alla posizione 16037 visibile su entrambi i filamenti (freccia rosa); con la freccia blu viene indicata una seconda eteroplasmia presente nella zona.

4.5 Variazioni polimorfiche a singolo nucleotide

Per ogni soggetto studiato è stata compilata una lista di polimorfismi, riferiti alla rRCS; in via preliminare, tuttavia, i polimorfismi sono stati classificati in base alla localizzazione sul genoma mitocondriale. Per ciascuno di questi è stato valutato se la variazione indotta fosse sinonima o meno e l'eventuale/presunta associazione a condizioni patologiche (per le abbreviazioni qui utilizzate, si invita il lettore a fare riferimento alle tabelle dettagliate allegate).

Di ciascuno dei polimorfismi osservati si è verificato il riscontro in letteratura, o la nuova descrizione.

Nei 40 campioni analizzati sono state individuate 917 variazioni polimorfiche a singolo nucleotide (includendo le due eteroplasmie) (Tab. IV.6 e fig. 4.16).

Il 97.7% dei polimorfismi (N=896) sono transizioni (568 purina-purina e 328 pirimidina-pirimidina) e il restante 2.3% sono transversioni (N=21). Il risultato è compatibile con quello atteso ed è sintetizzato nella Tabella IV.6. Infatti è stato dimostrato che la polimerasi γ possiede la funzione esonucleasica $3' \rightarrow 5'$ (correttore delle bozze o *proof reading*) che rende i processi replicativi altamente fedeli (l'elevato tasso di mutazione del mtDNA sembra derivare principalmente da uno sbilanciamento nella disponibilità di dNTPs). Quindi ci si aspetta che il tasso di mutazione permanga alto per le transizioni (dovute al tautomerismo, Introduzione) ma sia molto basso per le transversioni (Kaguni, 2004).

Polimorfismi TOTALI = 915							
Transizioni = 896				Transversioni = 21			
Purina-Purina = 568		Pirimid.-Pirimid. = 328		Purina-Pirimid. = 6		Pirimid.-Purina = 15	
A→G	G→A	C→T	T→C	A→T = 2	A→C = 3	T→A = 2	T→G = 2
424	142	149	179	G→T = 0	G→C = 1	C→A =10	C→G = 1

Tab.IV.6 Schema riassuntivo delle transversioni e delle transizioni osservate. La direzione delle variazioni si riferisce al filamento L.

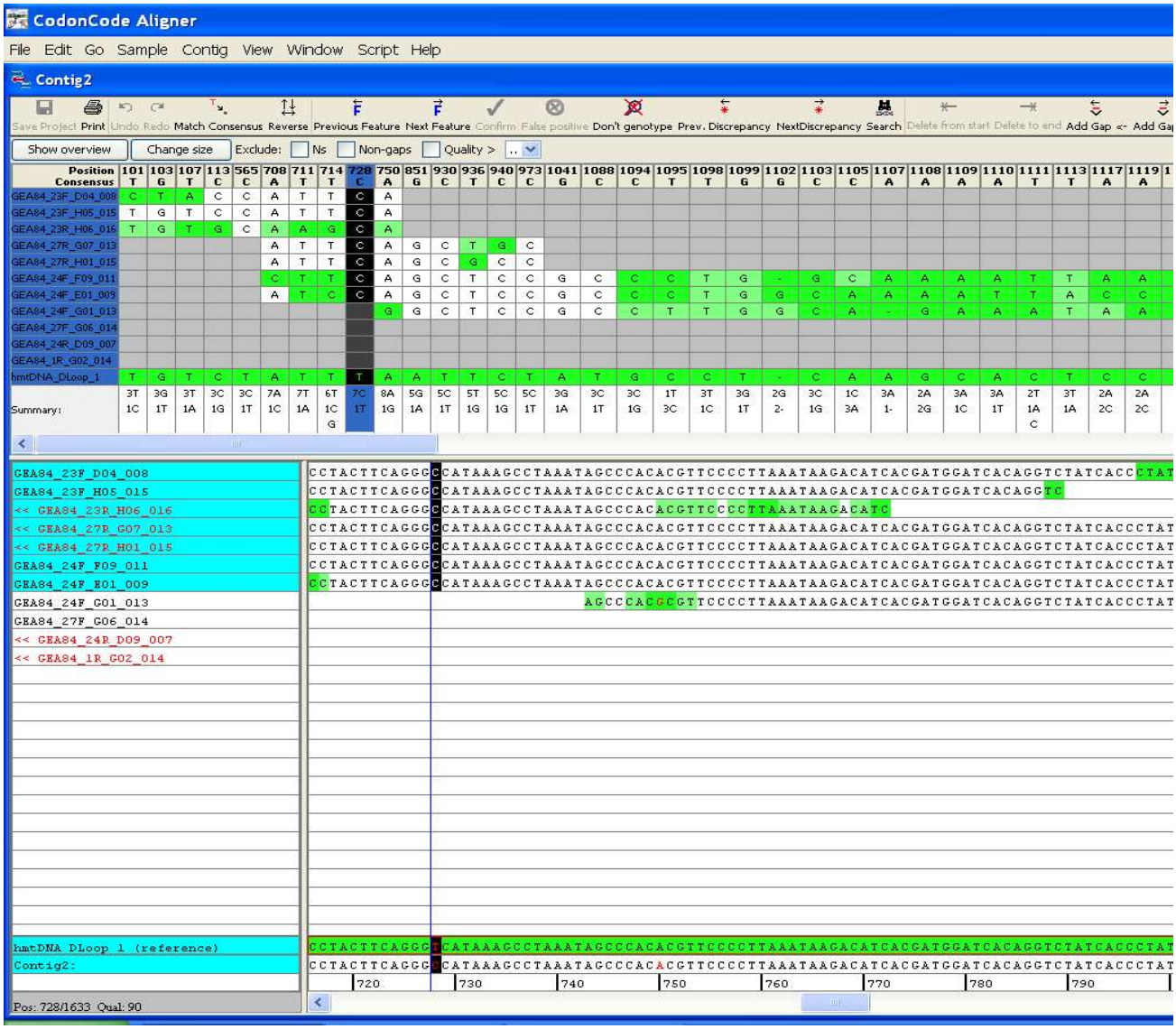


Fig. 4.16 Esempio di lettura dei polimorfismi. In basso sono visibili la sequenza di riferimento (HmtDNA D-loop), relativa alla regione di controllo, ed il consenso. Le differenze vengono riportate come discrepanze, che per praticità sono indicate in colore rosso, in corrispondenza della posizione. E' possibile visualizzare il relativo tracciato elettroforetico (vedere figure precedenti)

4.5.1 Polimorfismi nei geni dei tRNA (Tab. IV.7)

Molte mutazioni del DNA mitocondriale, definite come patologiche, sono localizzate nei tRna, specie nella porzione delle braccia (McFarland et al, 2004). Dal momento che le dimensioni dei tRna sono piuttosto simili, oscillando solo tra 65 e 75bp, vi sono delle peculiarità nella variabilità nei geni dei tRna. In questo studio sono stati individuati complessivamente 14 polimorfismi, da transizioni. Il tRna più polimorfico è risultato essere quello della leucina, presente nel 25% dei campioni, nel quale si ha variazione a livello della regione variabile della struttura a trifoglio. La mutazione in questione, la A12308G, è stata associata alla CPEO, all'Ictus, alla cardiomiopatia, al tumore renale e prostatico e ad alterato pH cellulare (www.mitomap.org). Seguono il $tRna^{Arg}$ (T10463C) ed il $tRna^{Thr}$ (G15928A), copresenti nel 10% dei campioni, con mutazioni a livello del braccio accettore. Nel caso dell'arginina vi è una sospetta variabilità fenotipica, nel caso della treonina è stata descritta un'associazione con la sclerosi multipla (Mayr-Wohlfart et al, 1996). Ancora, un altro polimorfismo a carico del $tRna^{Arg}$ (G10427A) è stato evidenziato in un altro soggetto, in posizione variabile, ed altri 3 polimorfismi a carico del $tRna^{Thr}$ sono stati evidenziati in altri 5 soggetti (ansa D e braccio accettore). Relativamente alla mutazione dell'ansa D (C15904T) vi è solo il sospetto di possibili variazioni fenotipiche, mentre per quelle che riguardano il braccio accettore sono state descritte associazioni con la LIMM (A15924G: Brown et al, 1992; Yoon et al, 1991) e con la sclerosi multipla ed aumento di penetranza della sordità associata alla mutazione 1555 (G15927A: Mayr-Wohlfart et al, 1996; Chen et al, 2008; Wang et al, 2008; Lu

et al, 2010). Il dato riguardante la treonina è in linea con altre osservazioni, riferenti un numero di mutazioni prevalente a livello del braccio accettore piuttosto che delle anse (Kivisild et al, 2006; Ruiz-Pesini et Wallace, 2006). Altre mutazioni associate a patologie sono state individuate nel $tRna^{Gln}$ (T4336C) relativamente all'ADPD ed alla perdita dell'udito, anche in condizione di eteroplasmia e comunque ancora non ben definite, e nel $tRna^{Lys}$ (A8296G), relativamente a DMDF, MERRF, HCM ed epilessia, anche in condizione eteroplasmica (www.mitomap.org). Un polimorfismo riscontrato in un solo individuo non è mai stato descritto in letteratura e coinvolge il tRNA per la valina ($tRNA^{Val}$); esso si colloca nell'ansa dell'anticodone, ma non fa parte di quest'ultimo. Un secondo, che riguarda il tRNA per l'acido aspartico ($tRNA^{Asp}$) e che parimenti è stato visto in un solo individuo, è inserito in mtDB ed ha 7 hits in Gene Bank.

Nelle figg 4.17 e 4.18 vengono mostrate la strutture secondarie dei due tRNA (http://mamit-trna.u-strasbg.fr/).

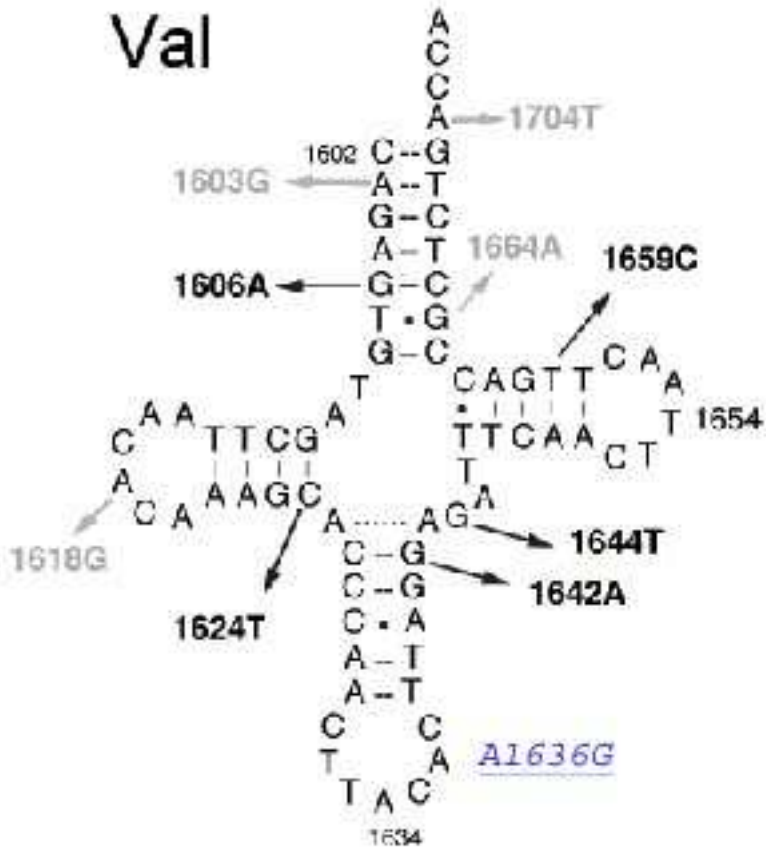


Fig. 4.17 Struttura secondaria del tRNA per la valina (Val). In grassetto sono evidenziate le mutazioni patogeniche, in grigio i semplici polimorfismi. Viene riportato il nuovo polimorfismo.tRNA^{Val}, situato nel braccio dell'anticodone.

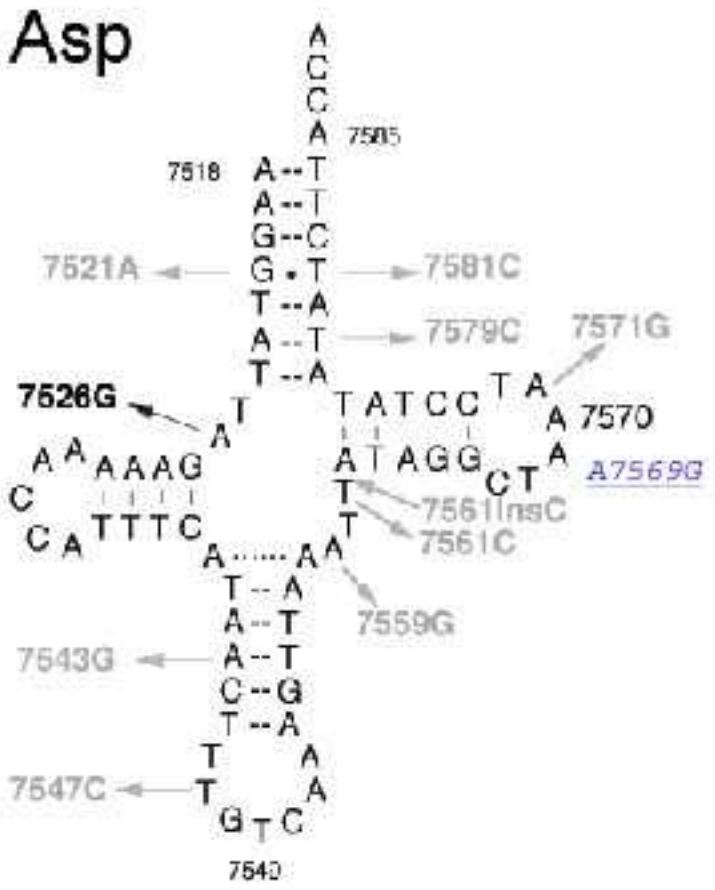


Fig. 4.18 Struttura secondaria del tRNA per l'acido aspartico (Asp). In grassetto sono evidenziate le mutazioni patogeniche, in grigio i semplici polimorfismi. Viene riportato il nuovo polimorfismo $tRNA^{Asp}$, situato nella zona dell'ansa T.

Nessuna delle due variazioni coinvolge le basi che appaiandosi permettono la formazione della struttura a trifoglio. La variazione del gene $tRNA^{Val}$ (A1636G) interessa il braccio dell'anticodone, più precisamente un nucleotide altamente conservato nei $tRNA^{Val}$ mitocondriali dei mammiferi: in tale posizione, infatti, è stata osservata sempre una purina (percentuale di conservazione 100%, Helm et al, 2000). La variazione del gene $tRna^{Asp}$ (A7569G) mappa nell'ansa T in una posizione che non è coinvolta nella formazione della struttura terziaria e non è altamente conservata (percentuale di conservazione inferiore al 50%, Helm et al, 2000).

4.5.2 Polimorfismi nei geni degli Rrna (Tab. IV.8)

Sono state identificate 30 variazioni totali, di cui 27 sono transizioni e 3 transversioni.

Per quanto riguarda l'rRna 12S, 4 mutazioni (G709A, A750G, G930Ae A1438G) sono state associate a sospetta variazione fenotipica. Vale la pena sottolineare che la A750G è presente nel 97.5% del campione esaminato e nel 98% dei campioni presenti in Gene Bank; parimenti, la A1438G è presente nel 92.5% del campione e nel 96% dei campioni presenti in Gene Bank. I relativi potenziali patogenici sono pertanto trascurabili. Una probabile associazione con la LVNC è stata invece descritta per la mutazione T921C (Tang et al, 2010), con la sordità per la mutazione A827G (www.mitomap.org), anche se i reports sono contrastanti. Sono infine presenti delle varianti non pubblicate relative alle mutazioni T794A e G1007A (vedere tabella).

Per quanto riguarda l'rRna 16S, 4 mutazioni (G1719A, A1811G, G1888A e A2706G) sono state associate a sospetta variazione fenotipica. Anche in questo caso vi è una mutazione, la A2706G, rappresentata nel 55% dei campioni in esame e presente nell'84% delle sequenze depositate in Gene Bank. Sono anche presenti delle varianti non pubblicate a carico delle mutazioni T1700C, C1703T E T3197C (vedere tabella). Tra i polimorfismi dell'rRrna 16S se ne annoverano 4, descritti in mtDB e con riscontri in Gene Bank: sono A1692T, A1943G, T2010C e T2159C. Un quinto polimorfismo non è inserito nei database tradizionali, ma ha un unico riscontro in

Gene Bank è il C2911T. Complessivamente, nel caso dell'rRNA 12S sono state individuate 11 mutazioni nelle anse della molecola (79%) e solo 3 nel braccio (21%). Parimenti, nell' rRrna 16S sono presenti 15 mutazioni nelle anse (94%) ed una sola nel braccio (6%). Questi dati sono in linea con gli studi effettuati su larga scala, a livello di strutture secondarie dei rispettivi rRna (Pereira et al, 2009).

4.5.3 Polimorfismi nei geni codificanti le subunità dei complessi enzimatici coinvolti nella fosforilazione ossidativa

I polimorfismi che rientrano in questa categoria sono numerosi. Poiché alcuni campioni possiedono due o più variazioni sullo stesso gene, è possibile ipotizzare che alcune di esse siano compensative.

4.5.3.1 Polimorfismi nei geni codificanti le subunità della citocromo c ossidasi (Tab. IV.9)

In totale sono stati contati 46 polimorfismi, tutti generati da transizioni, di cui 1 completamente nuovo, 1 nuovo come condizione eteroplasmica, con un solo altro report in Gene Bank, ed 8 variamente descritti. Il rapporto tra mutazioni sinonime (39) e non sinonime (7) è di circa 5.6 : 1.

Tra le mutazioni non sinonime (vedere tab), la T6253C è stata descritta in associazione a tumore prostatico (Brandon et al, 2006; Petros et al, 2005), mentre la T7270C è stata associata ad una sospetta variazione fenotipica, da confermare.

Per quanto riguarda le mutazioni sinonime, 5 di queste (A6146G, A7963G, G8251A, C9335T, G9477A) sono state inserite in MITOMAP nella “raccolta delle varianti non pubblicate”**,** riferentisi a soggetti con fenotipo patologico, nei quali il tipo di relazione di questo con i suddetti polimorfismi non è tuttora chiarita (vedere tabella). Tra le variazioni sinonime è stato individuato un polimorfismo avente un solo riscontro in GeneBank e qui descritto in condizione eteroplasmica (A6800G), a carico della Valina, ed uno di nuovissimo riscontro (T6512C), che interessa l’Alanina; questi ultimi due polimorfismi ricadono all’interno del gene COI.

4.5.3.2 Polimorfismi nei geni codificanti le subunità 6 e 8 dell’ATP sintasi (Tab. IV.10)

Nei due geni mitocondriali codificanti le subunità dell’ATP sintasi sono stati individuati 24 polimorfismi, di cui uno derivante da transversione (vedere oltre); un polimorfismo cade nella zona di sovrapposizione tra i geni ATP6 ed ATP8, pertanto viene conteggiato due volte. Dei 24 polimorfismi osservati, tre sono stati descritti nel mtDB, uno è rilevabile solo in Gene Bank, dove è presente in altri sei casi, ed un altro è di nuovissima descrizione. Contrariamente agli altri geni codificanti proteine, in questo caso il rapporto tra mutazioni sinonime (9) e non sinonime (15) è sbilanciato verso queste ultime, essendo di circa 1 : 1.67.

Andando a vedere nel dettaglio le mutazioni non sinonime, la A8531G, che cade all’interno della zona di sovrapposizione tra i due geni, non risulta associata a condizioni patologiche, e così pure la T8639C, riscontrabile in 6 sequenze depositate

in Gene Bank; per quanto riguarda la A9115G, si osservano due hits in Gene Bank, non correlati a fenotipi patologici, mentre, al contrario, la A8836G è stata strettamente associata alla LHON (Abu-Amero et Bosley, 2006). Di altre 5 variazioni polimorfiche (A8701G, T8705C, A8860G, A9007G, G9055A) è stata descritta una probabile associazione con fenotipi patologici, tuttavia da approfondire; vale la pena osservare che la variazione A8860G è presente in tutti i campioni esaminati, nonché nel 99.6% dei campioni depositati in Gene Bank: il suo potenziale patogenico è da ritenersi, pertanto, trascurabile. Situazione del tutto particolare riguarda la transversione T8945G, che non è mai stata descritta in alcun data base né individuata nei vari studi. Essa fa sì che la tripletta codificante per la Metionina, aminoacido neutro polare, si trasformi in un codone di stop. Il nucleotide cade specificatamente all'interno del gene ATP6, mentre i cinque casi simili, descritti in Gene Bank, riguardano mutazioni di stop a carico del Triptofano nelle posizioni 4720 e 5185, nel gene ND2, 11403 in ND4, la Metionina in 10657G, nel gene ND4L, ed infine la Lisina, in 15606, nel gene codificante il Citocromo B. Tutti questi casi sono stati evidenziati in pazienti affetti da tumore della tiroide. Stando a queste osservazioni ed ai dati visionabili in Gene Bank, relativi anche alla situazione opposta, cioè di codoni di stop che mutano verso aminoacidi codificanti (non descritta nel presente studio), si può osservare che mentre l'estensione di una proteina di alcuni aminoacidi, è di riscontro nella popolazione neutrale, l'accorciamento delle proteine è stato osservato unicamente nei casi patologici (Pereira et al, 2009). Nel caso specifico di questo

studio non si dispone di notizie circa lo stato clinico del soggetto portatore di tale mutazione. Sicuramente, però, se ne può dedurre l'elevato potenziale patogenico.

Tra le variazioni sinonime, la T8419C è stata descritta in mtDB ed individuata in 10 sequenze depositate in Gene Bank; anche in questo caso vi è un polimorfismo classificato come "variante non pubblicata", T8598C, il cui significato è da chiarire.

4.5.3.3 Polimorfismi che coinvolgono il gene codificante il citocromo b **(Tab. IV.11)**

In totale sono stati identificati 24 polimorfismi, uno dei quali descritto in mtDB e con altri riscontri in Gene Bank; due di questi polimorfismi sono transversioni. Il rapporto tra mutazioni sinonime (14) e non sinonime (10) è di 1.4 : 1.

Alla categoria delle mutazioni non sinonime appartengono: la C14766T, peraltro rappresentata nel 50% dei campioni del presente studio e della quale si sospetta una non chiara associazione a fenotipo patologico; lo stesso discorso vale per la A15326G, presente nel 100% dei campioni. Altre quattro mutazioni, meno rappresentate nella popolazione esaminata -G15734A**,** A14793G, T14798C e C15452A- sono parimenti associate a sospetta variazione fenotipica, in corso di approfondimento. Candidata ad un'associazione con la LVNC è invece la transizione T15693C (Tange et al, 2010), ma gli studi in tal senso non sono definitivi. Ricade tra le mutazioni non sinonime anche la variante non pubblicata C15632T.

Per quanto riguarda le mutazioni sinonime, la C15211T, è stata osservata in un solo caso ed ha 8 riscontri in Gene Bank. Più particolare è il caso della G15043A, descritta in associazione alle Principali Sindromi Depressive, ma nel contesto di uno specifico aplotipo, peraltro assente nel presente studio (Rollins et al, 2009).

4.5.3.4 Polimorfismi nei geni che codificano per le subunità della NADH deidrogenasi (Tab. IV.12)

In questa categoria sono stati identificati 111 polimorfismi, di cui 5 nuovi, 2 descritti unicamente in Gene Bank ed 11 variamente descritti; complessivamente sono presenti centosette transizioni e quattro transversioni, 82 mutazioni sinonime e 29 mutazioni non sinonime, con un rapporto pari a 2.83 : 1.

All'interno del gene ND1 sono presenti due transversioni, 12 mutazioni sinonime e 4 non sinonime, con un rapporto di 3 : 1. Per quanto riguarda le mutazioni non sinonime, due di esse, la T3398C e la T4216C sono state descritte in associazione a fenotipi patologici; la prima, in particolare, alla DMDF e HCM/GDM/probabile LVNC, la seconda alla LHON ed insulino-resistenza (www.mitomap.org). Quest'ultima, peraltro, è stata pure descritta come variante non pubblicata in un soggetto con adenoma sporadico delle paratiroidi, nel quale la relazione causa-effetto del polimorfismo con la patologia è del tutto da chiarire. Di nuovissima descrizione è la transversione A4021T, nella quale viene comunqe mantenuta la neutralità e la polarità dell'aminoacido risultante (dalla treonina alla serina, entrambi polari e neutri). Per quanto riguarda le mutazioni sinonime, sono

state individuate la C3435T, che ha unicamente riscontro in Gene Bank con altri 7 hits, la C3804T e la A4227G, entrambe descritte in mtDB. Anche in questo caso sono presenti varianti non pubblicate, vale a dire la T3847C e la transversione C3921A (vedere tabella per dettagli).

Nel gene ND2 sono presenti 22 transizioni e nessuna transversione, 14 mutazioni sinonime ed 8 non sinonime (rapporto 1.75 : 1). Tra le mutazioni non sinonime, la T4561C e la C4843T sono sospettate associate ad una qualche variazione fenotipica, mentre la possibile patogenicità per la G5460A, anche in eteroplasmia, è stata ipotizzata, ma non completamente chiarita, relativamente alla AD e PD (Mitchell et al, 2006; Tanaka et al, 2010). Altra associazione patologica, questa volta con la LHON, l'Insulino resistenza, la AMD e la NRTI-PN riguarda la mutazione A4917G (www.mitomap.org). Di nuovissima descrizione è la C5424T, nella quale un'istidina, aminoacido polare basico idrofilico, viene sostituita da una tirosina – aminoacido polare acido idrofilico; non vi è riscontro in letteratura né a livello di fenotipo individuale del soggetto interessato, circa l'eventuale effetto patogenico della mutazione. Tra le mutazioni sinonime, la A5120G ha riscontro in mtDB e Gene Bank, mentre la C5345T, a carico della fenilalanina, è di nuovissima descrizione.

Nel gene ND3 sono presenti 4 transizioni, nessuna transversione ed un rapporto tra mutazioni sinonime (2) e non sinonime (2) pari ad 1 : 1. Delle due mutazioni non sinonime ve ne è una potenzialmente associata a variazione fenotipica (A10398G),

mentre delle due sinonime, la T10248C è stata descritta in mtDB ed ha dei riscontri in Gene Bank.

Nel gene ND4L sono state riscontrate 3 transizioni e nessuna transversione, il rapporto tra mutazioni sinonime (2) e non sinonime (1) è pari a 2 : 1, la mutazione non sinonima, A10398G, è potenzialmente associata a variazione fenotipica, mentre la mutazione sinonima T10493C è inclusa in mtDB e descritta in Gene Bank.

Per quanto riguarda il gene ND4, sono presenti 17 transizioni e nessuna transversione, 15 mutazioni sinonime e 2 non sinonime, con un rapporto pari a 7.5 : 1. Tra le mutazioni sinonime, la T11009C è inserita i mtDB e descritta in Gene Bamk; particolare è invece il caso della A11467G, riscontrata nel 25% dei campioni esaminati e presumibilmente associata ad alterazioni del pH cerebrale post-mortem in soggetti affetti dalle Principali Sindromi Depressive (Rollins et al, 2009). Cinque mutazioni sinonime, infine, appartengono alle varianti non pubblicate, descritte in altrettanti soggetti che contestualmente presentavano determinati fenotipi patologici (vedere tabella): sono la T10810C, la T10873C, la C11332T, la A11362G e la C11674T. Infine, la C11356T, a carico della tirosina e di nuovissima descrizione, è stata individuata in un solo soggetto.

Nel gene ND5 sono presenti 38 mutazioni, tutti generati da transizioni, con un rapporto tra mutazioni sinonime (27) e non sinonime (11) pari a 2.45 : 1. Per quanto riguarda le mutazioni non sinonime, vi è una sospetta associazione fenotipica con le mutazioni A13105G e la C13934T, mentre una più chiara associazione patologica

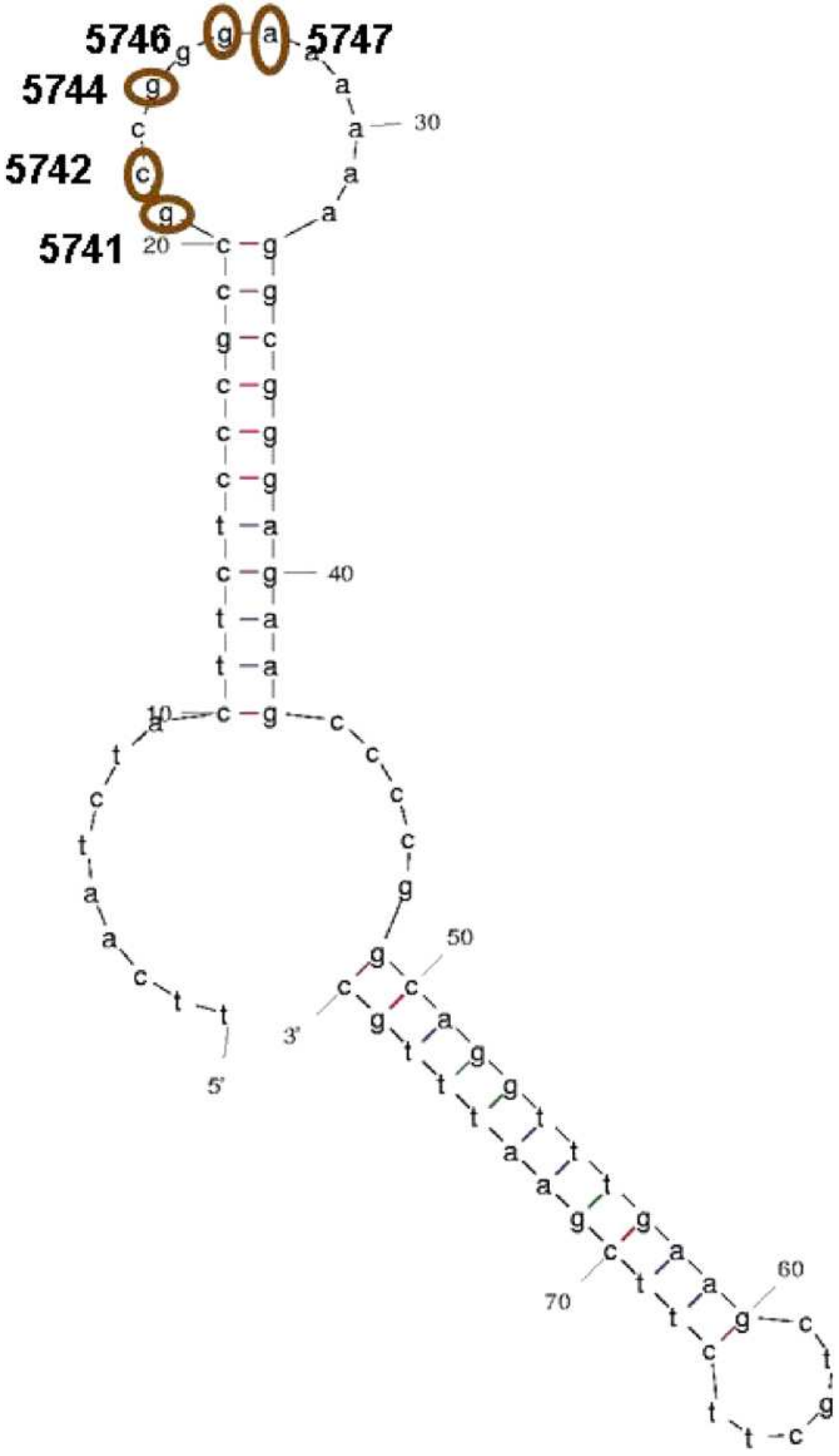
con la LHON è stata descritta per la G13708A (www.mitomap.org). Tra le mutazioni sinonime, è di rilevante osservazione la G12372A che, al pari della A11467G, sopra descritta, è stata riscontrata nel medesimo 25% dei soggetti presenti in questo studio, con le medesime possibilità di associazione alle Principali Sinddromi Depressive (Rollins et al, 2009). Anche in questo caso sono state descritte mutazioni incluse nel database delle varianti non pubblicate (vedere tabella per dettagli), che sono: la C12705T e la A12822G. Altre due mutazioni, la A12729G e la C12745T sono state descritte in mtDB e parimenti in Gene BanK, per entrambe esiste una sola descrizione in entrambi i database. La A13470G, descritta in entrambi i database, è invece più frequente (vedere Tabella). Infine, la C13785T è stata riscontrata unicamente in Gene Bnk, insieme ad altri quattro casi, mentre la A13866G, a carico della lisina ed individuata in un solo soggetto, è di nuovissima descrizione.

Nel gene ND6 sono state individuate 11 mutazioni, di cui una data da tranversione, ed un rapporto tra mutazioni sinonime (10) e non sinonime (1) pari a 10 : 1. La mutazione non sinonima A14193G è stata descritta in mtDB ed ha vari riscontri in Gene Bank, e così pure quella sinonima A14548G. Tra le mutazioni sinonime, sono presenti alcune descritte nel database delle varianti non pubblicate (vedere Tabella), che sono: la A14179G, la C14284T e la T14470C. Di quest'ultima è stata riscontrata pure la transversione T → A, a carico di un soggetto.

4.5.4 Polimorfismi nelle regioni non codificanti (Tab. IV.13)

In totale sono stati osservati 87 polimorfismi distribuiti nel D-loop, di cui 82 sono transizioni e 5 sono transversioni, con un rapporto pari a 16.4 : 1. Altri 2 polimorfismi, dati da transizioni, sono localizzati in tratti di sequenza non codificante, al di fuori del D-loop medesimo; in particolare, uno mappa nel tratto non codificante NC3 (in totale queste piccole zone sono 10), costituito da sette nucleotidi e posto tra il $tRNA^{Trp}$ e il $tRNA^{Ala}$. Il secondo mappa all'interno dell'origine di replicazione del filamento L costituita da 31 nucleotidi non sovrapponentisi ad altri geni, posti tra il $tRNA^{Asn}$ e il $tRNA^{Cys}$ (5730-5760), ed altri 9, iniziali, sovrapposti al $tRNA^{Asn}$ (5721-5729), più 38 finali che si sovrappongono al $tRNA^{Cys}$ (5761-5798). Il polimorfismo osservato (G5744A) è localizzato esattamente nella metà dell'O_L non sovrapponentesi agli altri geni; tale dato può essere importante per future speculazioni ed ipotesi circa la funzionalità e la struttura dell'O_L medesima. Il confronto con una struttura secondaria, calcolata mediante il software "mfold" (Zuker, 2003) localizza tale variante all'interno di una singola ansa predetta. Tale dato ha altri quattro riscontri in Gene Bank (Pereira et al, 2009). Nella fig. 4.19 viene indicata la variabilità all'interno della regione dell'Origine di Replicazione del Filamento L, paragonata alla dedotta struttura secondaria del DNA. Sono indicate le posizioni polimorfiche osservate a tale livello.

Fig. 4.19: Struttura secondaria di OL. (vedere in basso)


5746
5747
5744
5742
5741
5'
3'
10
20
30
40
50
60
70

Tab. IV.7. Polimorfismi nei geni dei tRNA

Variazione	tRNA	Regione interessata	Patologie**	Osservazioni***	Campioni	Totale
A1636G	V (valina)	**Ansa contenente l'anticodone**			109	1
T4336C	Q (glutammina)	**Braccio accettore**	**ADPD/H L - PhenoΔ**	**SNP per H5a**	**82**	1
T4452C	M (metionina)	**Ansa T**			111	1
T5528C	W (triptofano)	**Ansa D**			70	1
T5806C	C (cisteina)	**Altra**			61	1
A7569G	D (acido aspartico)	**Ansa T**		**Singola; 4 hits in GeneBank°**	79	1
A8296G	K (lisina)	**Braccio accettore**	**DMDF/ MERRF/ HCM/ Epilepsy**		68	1
G10427A	R (arginina)	**Altra**			69	1
T10463C	R (arginina)	**Braccio accettore**	**PhenoΔ**	**SNP per T**	**61,64,107,108**	4
A12308G	L (leucina)	**Regione variabile**	**CPEO/ICTUS/CM/Tumore renale e prostatico/Alterato pH cellulare - PhenoΔ**	**Già descritta come caratterizzante tutti gli Hg U e K**	**28,62,68,72,79,80,84,87,105,111**	10
C15904T	T (treonina)	**Ansa D**	**PhenoΔ**	**SNP già descritto per V**	**44,96**	2
A15924G	T (treonina)	**Braccio accettore**	**LIMM - PhenoΔ**	**SNP per U5a1a1**	**28,68**	2
G15927A	T (treonina)	**Braccio accettore**	**MS ed aumentata penetranza della DEAF 1555**		76	1

Variazione	tRNA	Regione interessata	Patologie**	Osservazioni***	Campioni	Totale
G15928A	T (treonina)	**Braccio accettore**	**MS - PhenoΔ**	**SNP per T**	**61,64,107,108**	4

Tab. IV.7 Polimorfismi osservati nei tRNA: sono evidenziati in giallo i polimorfismi assolutamente nuovi, mancanti di riscontro in tutti i database di riferimento; in verde i polimorfismi nuovi, non descritti in MITOMAP ed in mtDB, ma inseriti in Gene Bank; in azzurro quelli non inseriti in MITOMAP, ma osservati in casi sporadici e depositati in GeneBank e/o semplicemente riportati come osservazione (elencati nel database mtDB); in alcuni casi si è reso opportuno aggiornare le informazioni con il confronto diretto con i dati presenti in Gene Bank, nel qual caso sono state create delle postille (cerchietti evidenziati in grigio), riferendosi espressamente "alla data attuale". Sono sottolineati i polimorfismi selettivi per gli aplogruppi (vedere oltre), indicati nell'albero filogenetico di riferimento (PhiloTree), mentre non vengono sottolineati quelli che nei relativi soggetti appaiono come "mutazioni private"; si indicano in corsivo i soggetti nei quali il polimorfismo selettivo per l'aplogruppo appare come "mutazione privata". In blu sono indicati i polimorfismi "storici", definiti da tempo in letteratura, in fucsia quelli di più recente definizione ed altre caratteristiche. I nomi dei campioni sono indicati solo con il numero omettendo la serie (GEA); nell'ultima colonna è indicato il totale di soggetti che presentano il polimorfismo.

***Sinonima/Non sinonima**

****Patologie descritte in associazione al polimorfismo. Per ulteriori dettagli circa le condizioni patologiche indicate, si rimanda il lettore al sito web www.mitomap.org**

*****In questa colonna vengono riportati: a) fenotipi patologici singoli, osservati contestualmente alla presenza del polimorfismo. Allo studio in vista dell'inserimento nel data base precedente (**); b) caratteristiche di un dato polimorfismo a livello forense, riscontro nei data base di riferimento e così via.**

Abbreviazioni: **PhenoΔ =** Potenziale variazione fenotipica (cioè sospetta)**; ADPD** = Morbo di Alzheimer e Morbo di Parkinson**; HL =** perdita dell'udito**; DMDF** = Diabete Mellito e Sordità**; MERRF** = Epilessia Mioclonica e Fibre Rosse muscolari sfilacciate (Ragged-Red Fibers)**; HCM** = Cardio Miopatia Ipertrofica**; CPEO** = Oftalmoplegia Esterna Progressiva Cronica**; CM = Cardiomiopatia; LIMM** = Miopatia Mitocondriale Infantile Letale**; MS** = Sclerosi Multipla**; DEAF 1555** = Sordità associata alla variazione in 1555**.**

N.U. = Non Utile; **Hg =** aplogruppo; **MM =** Mutation Motif **Alla data attuale**: °7hits.

Tab. IV.8. Polimorfismi nei geni degli rRna

Variazione	rRNA	Regione interessata	Patologie**	Osservazioni***	Campioni	Totale
C679A	12S	**Ansa**			60	1
G709A	12S	**Ansa**	**PhenoΔ**	**SNP per H13a2, T, N2; privata in 2 casi**	**21,61,64,*80*,85,*104*,107,108**	8
A750G	12S	**Ansa**	**PhenoΔ**	**N.U.**	**Tutti tranne 83**	39
T794A	12S	**Ansa**		**Sordità (Africa); SNP per U6a7a1**	**87**	1
A827G	12S	**Ansa**	**DEAF**		80	1
T921C	12S	**Ansa**	**LVNC (probabile)**	**SNP per L3d1-5**	**63**	1
G930A	12S	**Ansa**	**PhenoΔ**	**SNP per T2b**	**107,108**	2
G1007A	12S	**Braccio**		**Adenoma sporadico delle paratiroidi (*USA*)**	109	1
A1008G	12S	**Ansa**		**SNP per H13a2a**	**21**	1
A1117G	12S	**Ansa**			57	1
T1193C	12S	**Ansa**		**SNP per U6a7a1**	**87**	1
T1243C	12S	**Braccio**		**SNP per W**	**85**	1
A1438G	12S	**Braccio**	**PhenoΔ**	**Già descritta come (assente) caratterizzante H2**	**Tutti tranne 21,83,98**	37

Variazione	Gene	Regione interessata	Patologie**	Osservazioni***	Campioni	Totale
G1598A	12S	**Ansa**		**SNP per N1b**	**70**	1
A1692T	16S	**Ansa**		**Singola; 7 hits in GeneBank°; SNP per U6a7**	**87**	1
T1700C	16S	**Ansa**		**Adenoma sporadico delle paratiroidi (*USA*) e tumore della bocca (Indiano Asiatico); SNP per U5a1a**	**28,62,68**	3
C1703T	16S	**Ansa**		**Ritardo mentale e tumore della mammella (Caucasici); SNP per N1b**	**70**	1
G1719A	16S	**Ansa**	**PhenoΔ**	**SNP per N1'5**	**59,70**	2
A1811G	16S	**Ansa**	**PhenoΔ**	**SNP per U2'3'4'7'8'**	**72,80,84,105,111**	5
G1888A	16S	**Ansa**	**PhenoΔ**	**SNP per T**	**61,64,107,108**	4
A1943G	16S	**Ansa**		**Singola; 1 hit in GeneBank°°**	72	1
T2010C	16S	**Braccio**		**Singola; 1 hit in GeneBank°°°**	64	1
T2159C	16S	**Ansa**		**Singola; 1 hit in GeneBank e 2 osservazioni non depositate°°°°**	71	1
C2259T	16S	**Ansa**		**SNP per H13a**	**21,60**	2
A2294G	16S	**Ansa**		**SNP per U3a**	**105,111**	2
C2639T	16S	**Ansa**		**SNP per N1b**	**70**	1

Variazione	Gene	Regione interessata	Patologie**	Osservazioni***	Campioni	Totale
A2706G	16S	**Ansa**	**PhenoΔ**	**Già descritta come (assente) caratterizzante H**	**28,44,59,61,62,63,64,67,68,70,72,79,80,84, 85,87,96,105,107,108,109,111**	22
C2911T	16S	**Ansa**		**Singola°°°°°**	79	1
G3010A	16S	**Ansa**		**Nel MM per H1 e SNP per J1 e U3a1**	**67,69,73,100,105,109**	6
T3197C	16S	**Ansa**		**Encefalopatia, cardiomiopatia (Mennonite) e tumore della bocca (Indiano Asiatico); SNP per U5**	**28,62,68,79**	4

Tab. IV.8 Polimorfismi degli rRNA. Valgono le annotazioni della Tabella IV.7.

Altre abbreviazioni: **LVNC** = Ridotta funzione sistolica del Ventricolo Sinistro**; DM** = Diabete Mellito**.**

Alla data attuale: °8 hits; °°5hits; °°°4 hits; °°°°8hits; °°°°°1 hit.

Tab. IV.9. Polimorfismi nei geni codificanti le subunità della citocromo *c* ossidasi

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A5951G	S	COI			71	1
T5999C	S	COI		**SNP per U4'9**	**80,84**	2
G6023A	S	COI		**SNP per U5a1f**	**79**	1
A6047G	S	COI		**SNP per U4**	**80,84**	2
A6146G	S	COI		**Infertilità (Africano/caucasico); SNP per U4c1**	**80**	1
T6152C	S	COI			111	1
T6221C	S	COI		**SNP per X**	**59**	1
T6253C	NS	COI	**PC**	**Nel MM per H15**	**57**	1
A6272G	S	COI		**Singola; 3 osservazioni non depositate; SNP per L3d1b1**	**63**	1
T6297C	S	COI			111	1
C6371T	S	COI		**SNP per X**	**59**	1
C6445T	NS	COI		**Singola; 1 hit in GeneBank**	64	1
T6512C	S	COI			87	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C6518T	S	COI		**SNP per U3a**	**105,111**	2
C6546T	NS	COI			72	1
A6599G	S	COI		**Singola; 4 hits in GeneBank°°**	72	1
A6629G	S	COI		**SNP per U5a1a2a**	**62**	1
A6647G	S	COI		**Singola; 2 hits in GeneBank°°°**	76	1
T6680C	S	COI		**SNP per L3d1**	**63**	1
T6719C	S	COI		**SNP per U5a1a2a**	**62**	1
A6791G	S	COI			59	1
A6800G#	S	COI		**Singola°°°°**	82	1
C7028T	S	COI		**Già descritta come (assente) caratterizzante H**	**28,44,59,61,62,63,64,67,68,70,72,79,80,84,85,87,96,105,107,108,109,111**	22
T7094C	S	COI		**Singola; 3 hits in GeneBank°°°°°**	71	1
T7270C	NS	COI	**PhenoΔ**		104	1
G7337A	S	COI		**SNP per H13a1a1**	**60**	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A7424G	S	COI		**SNP per L3d**	**63**	1
T7621C	S	COII			65	1
T7624C	S	COII			81	1
G7789A	S	COII		**SNP per J1d**	**109**	1
G7805A	NS	COII		**SNP per U6a**	**87**	1
A7963G	S	COII		**Adenoma sporadico delle paratiroidi (*USA*); SNP per J1d**	**109**	1
T8047C	S	COII		**Singola; 2 hits in GeneBank°°°°°°**	81	1
A8050G	S	COII		**Singola; 2 hits in GeneBank**	44	1
G8152A	S	COII			21	1
G8251A	S	COII		**Tumore della bocca (Indiano Asiatico-Bengali); SNP per N1b e W**	**70,85**	2
A8261G	NS	COII			70	1
G9266A	S	COIII		**SNP per U3a**	**105,111**	2
C9335T	S	COIII		**KSS (Caucasico/Africano); SNP per N1b1**	**70**	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A9336G	NS	COIII		**Singola; 2 hits in GeneBank e 1 osservazione non depositata°°°°°°°**	109	1
G9380A	S	COIII			44	1
G9477A	S	COIII		**Ridotta fertilità (*Ignoto*); SNP per U5**	**28,62,68,79**	4
T9540C	S	COIII		**SNP per L3 e già descritta (come assente) in N**	**63**	1
G9575A	S	COIII			21	1
T9615C	S	COIII			64	1
T9698C	S	COIII		**SNP per U8**	**72**	1

Tab. IV.9 Polimorfismi che coinvolgono i geni per la citocromo *c* ossidasi. Valgono le stesse considerazioni della Tabella IV.7.

In eteroplasmia in rapporto 1:1 (vedere paragrafi precedenti)

Altre abbreviazioni: **PC** = Tumore della prostata; **KSS** = Kearn-Sayre Syndrome.

Alla data attuale: ° = 10 hits; °° = 5 hits; °°° = 29 hits; °°°° = 1 hit; °°°°° = 7 hits; °°°°°° = 3 hits; °°°°°°° = 8 hits.

Tab. IV.10. Polimorfismi nei geni codificanti le subunità 6 e 8 dell'ATP sintasi

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C8410T	S	ATP8			70	1
T8419C	S	ATP8		**Singola; 5 hits in GeneBank e 1 osservazione non depositata°**	88	1
C8472T	NS	ATP8		**SNP per N1b**	**70**	1
T8473C	S	ATP8		**SNP per U6a7**	**87**	1
T8503C	S	ATP8			59	1
A8531G	NS/NS	ATP8/ATP6		**Singola; 1 hit in GeneBank°°**	88	1
T8598C	S	ATP6		**Epilessia mioclonica (Caucasico); SNP per H2b**	**98**	1
T8618C	NS	ATP6		**SNP per L3d**	**63**	1
T8639C	NS	ATP6		**Singola°°°**	65	1
G8697A	S	ATP6		**SNP per T**	**61,64,107,108**	4
A8701G	NS	ATP6	**PhenoΔ**	**SNP per L3 e già descritta (come assente) in N; privata in 1 caso**	**63,***71*	2
T8705C	NS	ATP6	**PhenoΔ**		28	1
T8763C	S	ATP6		**SNP per N1b1d**	**70**	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C8818T	S	ATP6		**SNP per U4a**	**84**	1
A8836G	NS	ATP6	**LHON**	**SNP per N1b**	**70**	1
A8860G	NS	ATP6	**Pheno∆**	**N.U.**	Tutti	40
T8945G	NS	ATP6	? (stop-codon)	**Singola**	21	1
G8994A	S	ATP6		**SNP per W**	**85**	1
A9007G	NS	ATP6	**Pheno∆**	**In 2 campioni T2b3**	107,108	2
G9055A	NS	ATP6	**Pheno∆**	**SNP per U8b**	**72**	1
T9070G	NS	ATP6		**SNP per U4c1**	**80**	1
A9115G	NS	ATP6		**Singola; 2 hits in GeneBank. In 2 campioni H33**	58,66	2
G9196A	NS	ATP6			76	1

Tab. IV.10 Polimorfismi che coinvolgono i geni per le subunità dell'ATP sintasi. Valgono le stesse considerazioni della Tabella IV.7; il polimorfismo A8531G coinvolge entrambi i geni, poiché sono parzialmente sovrapposti.

Altre abbreviazioni: **LHON** = Neuropatia Ottica di Leber Ereditaria.

Alla data attuale: ° = 10 hits; °° = 3 hits; °°° = 6hits.

Tab. IV.11. Polimorfismi che coinvolgono il gene codificante il citocromo b

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A14755G	S	CytB			59	1
C14766T	NS	CytB	**PhenoΔ**	**Già descritta come caratterizzante tutti gli Hg NON HV**	**28,59,61,62,63,64,67,68,70,72,79,80,84,85,87,105,107,108, 109,111**	20
A14793G	NS	CytB	**PhenoΔ**	**SNP per U5a**	**28,62,68,79**	4
T14798C	NS	CytB	**PhenoΔ**	**SNP per J1c**	**67**	1
G14831A	NS	CytB			72	1
C14866T	S	CytB		**SNP per U4c1**	**80**	1
C14872T	S	CytB		**SNP per H13**	**21,60**	2
G14905A	S	CytB		**SNP per T**	**61,64,107,108**	4
G15043A	S	CytB	**MDD-associata**	**SNP per U6a7; privata in un caso**	**87**,*109*	2
C15211T	S	CytB		**Singola; 1 osservazione non depositata°**	44	1
A15218G	S	CytB		**SNP per U5a1**	**28,62,68,79**	4
G15301A	S	CytB		**Già descritto come SNP per L3**	**63**	1
A15326G	NS	CytB	**PhenoΔ**	**N.U.**	Tutti	40

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C15452A	NS	CytB	**PhenoΔ**	**SNP per JT**	**61,64,67,107,108,109**	6
T15454C	S	CytB		**SNP per U3**	**105,111**	2
C15455T	NS	CytB			96	1
T15511C	S	CytB			82	1
T15530C	S	CytB		**SNP per U6a7**	**87**	1
A15607G	S	CytB		**SNP per T**	**61,64,107,108**	4
C15632T	NS	CytB		**Cardiomiopatia (Tunisino); SNP per U6a7**	**87**	1
T15693C	NS	CytB	**LVNC (probabile)**	**SNP per U4**	**80,84**	2
G15734A	NS	CytB	**PhenoΔ**		105	1
C15833T	S	CytB		**SNP per H5a1**	**82**	1
G15884C	NS	CytB		**SNP per W**	**85**	1

Tab.IV.11 Variazioni sul gene codificante il citocromo b. Valgono le stesse osservazioni della Tabella IV.7

Altre abbreviazioni: **MDD** = Principali Sindromi Depressive.

°8 hits alla data attuale.

Tab. IV.12. Polimorfismi nei geni che codificano per le subunità della NADH deidrogenasi

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C3342T	S	ND1		**SNP per H35**	**81**	1
A3348G	S	ND1		**SNP per U6**	**87**	1
T3398C	NS	ND1	**DMDF + HCM/GDM/probabile LVNC**		88	1
C3435T	S	ND1		**Singola°**	86	1
A3480G	S	ND1		**SNP per U8b**	**72**	1
A3505G	NS	ND1		**SNP per W**	**85**	1
T3540C	S	ND1			72	1
C3804T	S	ND1		**Singola; 1 hit in GeneBank**	82	1
T3847C	S	ND1		**NAION (Arabi); SNP per H15b**	**57**	1
G3918A	S	ND1			71	1
C3921A	S	ND1		**NAION (Arabo) e tumore della mammella (Caucasico); SNP per N1b**	**70**	1
A4021T	NS	ND1		**Singola**	72	1
G4113A	S	ND1			64	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A4164G	S	ND1			79	1
T4216C	NS	ND1	**LHON ed insulino resistenza**	**Adenoma sporadico delle paratiroidi (*USA*); SNP per R2'JT e H10a**	**61,64,67,88,107,108,109**	7
A4227G	S	ND1		**Singola; 2 hits in GeneBank°°**	81	1
T4561C	NS	ND2	**PhenoΔ**		59	1
G4580A	S	ND2		**SNP già descritto per V**	**44,96**	2
T4646C	S	ND2		**SNP per U4**	**80,84**	2
G4655A	S	ND2			67	1
T4703C	S	ND2		**SNP per U3a**	**105,111**	2
A4745G	S	ND2		**SNP per H13a1**	**60**	1
A4769G	S	ND2		**Già descritta come (assente) caratterizzante H2**	**Tutti tranne 83**	39
A4811G	S	ND2		**SNP per U4c1**	**80**	1
C4843T	NS	ND2	**PhenoΔ**	**In 2 campioni T2b3**	107,108	2
A4917G	NS	ND2	**LHON/Insulino Resistenza/AMD/NRTI-PN**	**SNP per T**	**61,64,107,108**	4

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C4960T	NS	ND2		**SNP per N1b**	**70**	1
G5046A	NS	ND2		**SNP per L3d1b ed N2**	**63,85**	2
A5120G	S	ND2		**Singola; 7 hits in GeneBank°°°; SNP perU6a7a**	**87**	1
G5147A	S	ND2		**SNP per L3d1b e T2b**	**63,107,108**	3
A5198G	S	ND2			67	1
G5231A	S	ND2			67	1
A5319G	NS	ND2		**SNP per U5a1a2a**	**62**	1
C5345T	S	ND2		**Singola**	61	1
C5424T	NS	ND2		**Singola**	111	1
G5460A	NS	ND2	**PhenoΔ/ AD/PD**	**SNP per W**	**85**	1
G5471A	S	ND2		**SNP per N1b e U6a7; privata in un caso**	*64*,**70,87**	3
T5495C	S	ND2		**SNP per U5a1a1**	**28,68**	2
A10086G	NS	ND3			72	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
T10238C	S	ND3		**SNP per N1**	**70**	1
T10248C	S	ND3		**Singola; 1 hit in GeneBank°°°°**	59	1
A10398G	NS	ND3	**PhenoΔ**	**SNP per L3 e per J**	**63,67,109**	3
T10493C	S	ND4L		**Singola; 3 hits in GeneBank°°°°°**	84	1
A10506G	S	ND4L		**SNP per U3a**	**105,111**	2
A10750G	NS	ND4L	**PhenoΔ**	**SNP per T2b3**	**107,108**	2
T10810C	S	ND4		**Tumore del polmone (Africano/Americano)**	62	1
T10873C	S	ND4		**Tumore della bocca (Indiano Asiatico); SNP per L3 e già descritta (come assente) in N**	**63**	1
T10907C	NS	ND4		**SNP per U4c**	**80**	1
A10978G	S	ND4			68	1
G10993A	S	ND4		**In 2 campioni T2b3**	107,108	2
T11009C	S	ND4		**Singola; 4 hits in GeneBank°°°°°°; SNP per U4c1a**	**80**	1
A11251G	S	ND4		**SNP per JT**	**61,64,67,107,108,109**	6

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C11332T	S	ND4		**Tumore della bocca (Indiano Asiatico); SNP per U4**	**80,84**	2
C11356T	S	ND4		**Singola**	61	1
A11362G	S	ND4		**NAION (Arabi) e KSS (Caucasico/Africano); SNP per N1b1**	**70**	1
G11447A	NS	ND4		**SNP per H33**	**58,66**	2
A11467G	S	ND4	**Alterato pH cerebrale**	**Già descritta come caratterizzante tutti gli Hg U e K**	**28,62,68,72,79,80,84,87,105,111**	10
C11674T	S	ND4		**Ritardo mentale (Caucasico); SNP per N2**	**85**	1
G11719A	S	ND4		**Già descritta come caratterizzante tutti gli Hg NON R0 (HV, V ed H)**	**28,59,61,62,63,64,67,68,70,72,79,80,84,85,87,105,107,108,109,111**	20
A11812G	S	ND4		**SNP per T2**	**64,107,108**	3
G11914A	S	ND4			104	1
A11947G	S	ND4		**SNP per W**	**85**	1
C12346T	NS	ND5		**SNP per U5a1a2; privata in un caso**	**62**,*67*	2
G12372A	S	ND5	**Alterato pH cerebrale**	**Già descritta come caratterizzante tutti gli Hg U e K**	**28,62,68,72,79,80,84,87,105,111**	10
C12403T	NS	ND5			59	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
T12408C	S	ND5			64	1
T12414C	S	ND5		**SNP per N2**	**85**	1
G12501A	S	ND5		**SNP per N1**	**70**	1
G12528A	S	ND5			109	1
A12612G	S	ND5		**SNP per J**	**67,109**	2
G12618A	S	ND5			81	1
C12633A	S	ND5		**SNP per T1**	**61**	1
A12662G	NS	ND5			21	1
C12705T	S	ND5		**Oftalmoparesi e ptosi oculare (Asiatico); SNP nei NON N**	**59,63,70,85**	4
A12729G	S	ND5		**Singola; 1 osservazione non depositata°°°°°°°**	67	1
C12745T	S	ND5		**Singola; 1 hit in GeneBank**	63	1
G12771A	S	ND5			72	1
C12813T	S	ND5			96	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
A12822G	S	ND5		**NAION (Arabi); SNP per N1b**	**70**	1
A12930G	S	ND5			111	1
A13105G	NS	ND5	**PhenoΔ**	**SNP in L3c'd'j; privata in 1 caso**	*21*,**63**	2
G13135A	NS	ND5			109	1
T13260C	S	ND5			64	1
A13269G	S	ND5			72	1
T13326C	S	ND5		**SNP per H13a1a1a**	**60**	1
G13368A	S	ND5		**SNP per T**	**61,64,107,108**	4
A13470G	S	ND5		**Singola; 4 hits in GeneBank°°°°°°°°**	62	1
T13617C	S	ND5		**SNP per U5**	**28,62,68,79**	4
C13680T	S	ND5		**SNP per H13a1a**	**60**	1
G13708A	NS	ND5	**LHON**	**SNP per J; privata in un caso**	*59*,**67,109**	3
G13711A	NS	ND5			88	1

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
G13759A	NS	ND5			73	1
C13785T	S	ND5		**Singola°°°°°°°°°**	76	1
A13866G	S	ND5		**Singola**	70	1
T13886C	NS	ND5		**SNP per L3d**	**63**	1
T13899C	S	ND5			59	1
C13934T	NS	ND5	**PhenoΔ**	**SNP per U3a**	**105,111**	2
A13966G	NS	ND5		**SNP per X**	**59**	1
A14037G	S	ND5			83	1
A14139G	S	ND5		**SNP per U3**	**105,111**	2
C14167T	S	ND6		**SNP per U8b**	**72**	1
A14179G	S	ND6		**Sindrome di Leigh; SNP per U6a**	**87**	1
A14193G	NS	ND6		**Singola; 1 osservazione non depositata°°°°°°°°°**	71	1
A14233G	S	ND6		**SNP per T2**	**64,107,108**	3

Variazione	S/NS*	Gene	Patologie**	Osservazioni***	Campioni	Totale
C14284T	S	ND6		**LHON; SNP per L3d**	**63**	1
G14323A	S	ND6			64	1
T14470C	S	ND6		**MELAS (3243) (Caucasico); SNP per X**	**59**	1
T14470A	S	ND6		**SNP per H10**	**88**	1
A14548G	S	ND6		**Singola; 2 hits in GeneBank e 1 osservazione non depositata°°°°°°°°°°° già SNP per H10a1**	**88**	1
A14605G	S	ND6			64	1
C14620T	S	ND6		**SNP per U4**	**80,84**	2

Tab.IV.12 Polimorfismi nei geni codificanti le subunità della NADH deidrogenasi: valgono le stesse considerazioni della Tabella IV.7

Altre abbreviazioni: **GDM** = Diabete gestazionale; **NAION** = Neuropatia Ottica Ischemica Anteriore Non Arteritica. **AMD** = Degenerazione aculare associata all'età; **NRTI-PN** = Neuropatia periferica da Inibitore della Transcrittasi Inversa Nucleosidica (farmaco antivirale usato nella terapia contro l'HIV).

Alla data attuale: °7 hits; °°3 hits; °°°8 hits; °°°°3 hits; °°°°°7 hits; °°°°°°11 hits; °°°°°°°1 hit; °°°°°°°°32 hits; °°°°°°°°°6 hits; °°°°°°°°°°11 hits; °°°°°°°°°°4 hits.

Tab. IV.13. Polimorfismi nelle regioni non codificanti

Variazione	Patologie**	Osservazioni***	Campioni	Totale
T55C		**Nel MM per H15**	**57**	1
T57G		**Nel MM per H15**	**57**	1
T72C		**Nel MM per HVO**	**44,96**	2
A73G		**Già descritta come caratterizzante tutti gli Hg NON R0 (HV, V ed H)**	**28,59,61,62,63,64,67,68,70,72,79,80,84,85,87,105,107,108,109,111**	20
A93G			83	1
T146C			60	1
C150T		**Tumore della bocca (Indiani Asiatici); nel MM per U3a e privata in un caso**	*72*,**105,111**	3
C151T		**Tumore della bocca (Indiani Asiatici Marati); In 2 campioni T2b3**	107,108	2
T152C		**Nel MM di L3d, H2b, J1d e SNP per N1b; privata in 9 casi**	*44*,*59*,**63**,*65*,*69*,**70**,*76*,*80*,*84*,**98**,*104*,**109**,*111*	13
C182T			21	1
G185A		**Nel MM per J1c2 e SNP per N1b1d**	**67,70**	2
A188G		**Nel MM per J1c2 e SNP per N1b1d**	**67,70**	2
A189G		**Nel MM per W3'5**	**85**	1

Variazione	Patologie**	Osservazioni***	Campioni	Totale
C194T		**Nel MM per W3'5; privata in 1 caso**	**85**,*96*	2
T195C	**SZ/DB**	**Tumore della bocca (Indiani Asiatici – Marati); nel MM per X2, U4c1, U4a2, W3'5 e SNP in U8b1 e H11'12**	**59,72,80,84,85,104**	6
T199C			82	1
A200G			61,104	2
T204C		**Nel MM per W3'5**	**85**	1
G207A		**Nel MM per W3'5**	**85**	1
A214G			109	1
G228A		**SNP per J1c2**	**67**	1
A263G		**N.U.**	Tutti	40
C295T		**Nel MM per J1c2 e J1d (J)**	**67,109**	2
T310C		**Nel MM per U4a2**	**84**	1
T455C		**Singola°**	68	1
C456T		**Nel MM per H5**	**82**	1

Variazione	Patologie**	Osservazioni***	Campioni	Totale
T460C			68	1
C462T		**Nel MM per J1c2 e SNP per J1d (J1)**	**67,109**	2
T466C			69	1
T477C		**Nel MM per H1c**	**73**	1
T485C			28	1
T489C		**Nel MM per J1c2 e J1d**	**67,109**	2
G499A		**Nel MM per U4c1 eU4a2**	**80,84**	2
A508G			105	1
A574C (573.1C)		**Tumore del seno (Asiatico); SNP per U5a1a2**	**62**	1
G5585A		**Oftalmoparesi e ptosi dell'occhio singolo (Asiatico)**	79	1
G5744A			73	1
A16037G#			104	1
C16069T		**Nel MM per J1c2 e J1d (J)**	**67,109**	2

Variazione	Patologie**	Osservazioni***	Campioni	Totale
T16092C		**Tumore della bocca (Indiano Asiatico – Marati)**	67	1
T16093C			71	1
C16114T		**Nel MM per H10a**	**88**	1
T16124C		**Nel MM per L3d**	**63**	1
T16126C		**Nel MM per J1c2, J1d, T1a, T2 e T2b (JT)**	**61,64,67,107,108,109**	6
G16129A			70,96	2
G16145A		**Nel MM per N1b**	**70**	1
A16163G		**Nel MM per T1a**	**61**	1
T16172C		**Tumore della bocca (Indiano Asiatico); nel MM per U6 e privata in tre casi**	*28,61,72*,**87**	4
C16176G		**Nel MM per N1b**	**70**	1
C16179T		**Nel MM per U4c1**	**80**	1
(A16183C)		**Non considerata nell'albero filogenetico**	72,81	2
C16186T		**Nel MM per T1a; privata in due casi**	**61**,*69,109*	3

Variazione	Patologie**	Osservazioni***	Campioni	Totale
C16188T			104	1
T16189C	**Diabete II/ cardiomiopatia e rischio di tumore endometriale**	**Nel MM di X2, T1a eU8b; SNP in H1b e privata in 1 caso**	**59,61,69,72,***81*	5
C16192T		**Tumore della bocca (Indiano Asiatico – Marati); SNP per U5**	**79**	1
C16193T		**Nel MM per J1d**	**109**	1
T16209C			98	1
A16215G			59	1
A16219G		**Nel MM per U6**	**87**	1
C16221T			28	1
C16223T		**Tumore della bocca (Indiano Asiatico); nel MM perL3d, N1b e W3'5; privata in 1 caso**	*59*,**63,70,85**	4
C16234T		**Tumore della bocca (Indiano Asiatico – Gujarati); nel MM per U8**	**72**	1
A16235G		**Tumore della bocca (Indiano Asiatico); nel MM per H2a2b**	**83**	1
C16239A			96	1
C16245T			65	1

Variazione	Patologie**	Osservazioni***	Campioni	Totale
C16256T		**Tumore della bocca (Indiano Asiatico); nel MM per U5a**	**28,62,68,79**	4
C16270T		**Tumore della bocca (Indiano Asiatico); nel MM per U5a e privata in due casi**	**62,68,79***,107,108*	5
C16278T		**Tumore della bocca (Indiano Asiatico - Marati); nel MM di X2; privata in un caso**	**59***,69*	2
A16284G			*73*	1
C16291T		**Nel MM di H2a2b; privata in un caso**	*70*,**83**	2
C16292T		**Tumore della bocca (Indiano Asiatico - Marati); nel MM per W3'5**	**85**	1
C16294T		**Nel MM per T1a, T2 e T2b (T); privata in 2 casi**	*59*,**61,64**,*79*,**107,108**	6
C16296T		**Nel MM per T2 e T2b**	**64,107,108**	3
C16297T			60	1
T16298C		**Nel MM per HVO; privata in 1 caso**	**44,***61*,**96**	3
A16300G	**Associato alla BD**	**SNP per J1d1**	**109**	1
T16304C		**Nel MM per H5 e T2b**	**82,107,108**	3
A16309G		**Tumore della bocca (Indiano Asiatico - Muslim)**	65	1

Variazione	Patologie**	Osservazioni***	Campioni	Totale
T16311C		**Tumore della bocca (Indiano Asiatico); nel MM per H2b e privata in tre casi**	*72,79,***98**,*105*	4
G16319A		**Tumore della bocca (Indiano Asiatico)**	21	1
C16320T		**Tumore della bocca (Indiano Asiatico - Marati)**	85	1
A16343G		**Nel MM per U3a**	**105,111**	2
C16355T		**Tumore della bocca (Indiano Asiatico)**	73	1
T16356C		**Tumore della bocca (Indiano Asiatico); nel MM per U4c1 e U4a2; privata in un caso**	**80,84,***105*	3
T16362C		**Tumore della bocca (Indiano Asiatico - Marati)**	100	1
T16368C			86	1
G16390A		**Nel MM per N1b e U3a; privata in un caso**	*59,***70,105,111**	4
A16399G		**Nel MM per U5a1; privata in1 caso**	**28**,*61*,**62,68,79**	5
(T16519C)		**Non considerata nell'albero filogenetico**	21,59,61,63,64,65,69,71,72,73,76, 80,81,84,85,86,88,100,104,105,107,108,111	23

Tab.IV.13 Polimorfismi osservati situati nelle regioni non codificanti del DNA mitocondriale. Valgono le stesse considerazioni della Tabella IV.7.

In verde è segnalato il polimorfismo della zona NC4 (descritto nel testo).

#In eteroplasmia in rapporto 1:1 (vedere paragrafi precedenti).

Altre abbreviazioni: **SZ** = Schizofrenia; **BD** = malattia dipolare. **Alla data attuale:** °1 hit

Per quanto riguarda i polimorfismi individuati nel D-loop, vi sono la mutazione T195C e la A16300G per le quali si hanno dei report di associazione ai Disordini Bipolari (Rollins et al, 2009); la T16189C è stata invece descritta in associazione al Diabete di tipo II, cardiomiopatia e rischio di tumore endometriale (www.mitomap.org). Anche per questa categoria di polimorfismi sono state descritte delle varianti non pubblicate relativamente a C150T, C151T, la già ricordata T195C, la A574C, la T16092C, la T16172C, C16192T, C16223T, C16234T, A16235G, C16256T, C16270T, C16278T, C16292T, A16309G, T16311C, G16319A, C16320T, C16355T, T16356C, T16362C (vedere tabella). Per quanto riguarda i polimorfismi inseriti nelle regioni non codificanti, poste al di fuori del D-loop, si segnala la variante non pubblicata relativa alla mutazione G5585A, in NC3. E' stato, infine, individuato un polimorfismo finora non descritto nei database tradizionali, ma che ha un unico riscontro in Gene Bank (in HVIII).

4.5.5 Analisi dei polimorfismi

I polimorfismi osservati sono stati analizzati globalmente per ricavarne dei dati statistici. Nella tabella IV.14 è riassunta la variabilità media riscontrata nei campioni in base alla zona di provenienza, calcolata rispetto alla sequenza di riferimento già citata. Posto che il numero di campioni analizzati è limitato, i dati indicano la mancanza di una relazione tra il numero di polimorfismi osservati e la zona di provenienza dei campioni.

Zona	Campione (n°di polimorfismi)	Media di polimorfismi osservati
NORD	GEA65 (12)	**22.1**
	GEA73 (13)	
	GEA76 (12)	
	GEA79 (30)	
	GEA82 (14)	
	GEA86 (9)	
	GEA107 (38)	
	GEA108 (38)	
	GEA111 (33)	
CENTRO	GEA28 (28)	**19.2**
	GEA44 (16)	
	GEA80 (35)	
	GEA81 (14)	
	GEA88 (15)	
	GEA98 (9)	
	GEA100 (9)	
	GEA104 (15)	
	GEA105 (32)	
SUD	GEA57 (11)	**26.0**
	GEA58 (8)	
	GEA59 (34)	
	GEA60 (15)	
	GEA61 (37)	
	GEA62 (30)	
	GEA63(32)	
	GEA64 (38)	
	GEA66 (8)	
	GEA67 (33)	
	GEA68 (29)	
	GEA69 (14)	
	GEA70 (41)	
	GEA71 (14)	
	GEA72 (36)	
	GEA109 (36)	
ISOLE	GEA21 (17)	**21.8**
	GEA83 (7)	
	GEA84 (29)	
	GEA85 (32)	
	GEA87 (29)	
	GEA96 (17)	

Tab. IV.14 Schema riassuntivo dei polimorfismi individuati in ogni campione: nell'ultima colonna è riportata la media dei polimorfismi riscontrati nelle 4 zone analizzate.

L’unica indicazione riguarda la variabilità nel Sud Italia, che sembra particolarmente elevata (come indicano anche i dati discussi in seguito).

Per valutare se le sequenze presentano una variabilità maggiore nelle zone non codificanti è stata confrontata la somma di quelli osservati nelle zone codificanti e non in relazione alla lunghezza (Tab. IV.15). I dati indicano una variabilità di circa 7 nucleotidi ogni 100 nelle zone non codificanti, contro i 2 nucleotidi ogni 100 delle zone codificanti. I dati confermano quindi che la regione non codificante, in particolare quella del D-Loop, presenta il maggiore tasso di variazione, non solo per quanto riguarda i polimorfismi di singoli nucleotidi, ma anche riguardo alle inserzioni, alle delezioni e alle eteroplasmie di lunghezza, osservate prevalentemente all’interno del D-loop stesso.

REGIONI	Lunghezza Tot. (L.T.)	Polimorfismi Osservati (P.O.)	RAPPORTO P.O./L.T.
Non codificanti	~1308bp	90	0.069
Codificanti	~15261bp	249	0.016

Tab. IV.15 Variabilità della sequenza di mtDNA nelle zone codificanti e non codificanti.

Un’analisi simile è stata effettuata sui polimorfismi che cadono nelle regioni codificanti (Tab. IV.16), suddividendoli in base al tipo di gene coinvolto.

REGIONE	Lunghezza Tot. (L.T.)	Polimorfismi Osservati (P.O.)	RAPPORTO P.O/L.T.
rRNA	~2612bp	30	0.011
tRNA	~1511bp	14	0.009
Subunità enzimatiche	~11138bp	205	0.018

Tab. IV.16 Variabilità del mtDNA nella regione codificante.

In questo caso, la variabilità nella regione dei tRNA e degli rRNA è circa uguale a 1 polimorfismo ogni 100 nucleotidi, mentre risulta quasi il doppio nella regione codificante le subunità enzimatiche coinvolte nella fosforilazione ossidativa. I dati suggeriscono quindi che le zone più conservate all'interno della molecola sono quelle codificanti gli rRNA e i tRNA, come atteso. Infatti le mutazioni all'interno di questi geni sono tra le più dannose osservate (vedi capitolo 1) e ci si aspetta che le variazioni che mappano in queste regioni siano selezionate negativamente.

Nella tabella IV.17 lo stesso procedimento è stato applicato ai geni codificanti le subunità enzimatiche coinvolte nella catena respiratoria. I risultati indicano che i geni codificanti le subunità I, II e III della citocromo *c* ossidasi sono i meno soggetti a mutazioni, al pari dei geni codificanti le subunità della NADH deidrogenasi. Al contrario, i geni dell'ATP sintasi presentano la variazione di sequenza più alta tra tutti i geni della regione codificante. Ciò conferma i risultati di studi precedenti (Ingman e Gyllesten, 2001; Mishmar et al, 2003), ma nel contempo è in qualche modo un dato sorprendente, dal momento che in particolare il gene ATP6 è stato dimostrato essere altamente conservato anche tra specie lontanamente correlate (Wallace et al, 1987; Saccone et al, 2000).

Un altro dato interessante, a tal proposito, riguarda il numero di varianti non sinonime vs quelle sinonime, nei geni codificanti, ed in particolare, in ATP6 ed ATP8: qui, nuovamente, si riscontra la percentuale più alta di variazioni non sinonime, che, unitamente all'elevato tasso di variazione polimorfica, appena ricordato, in ATP6 (non

riscontrabile in specie correlate), fa ipotizzare che la selezione naturale possa avere agito sulla variazione in queste regioni, in un qualche punto dell'evoluzione umana (Ingman e Gyllensten, 2007). Studi di popolazione hanno evidenziato che la variazione non sinonima in ATP6 è significativamente più abbondante nelle popolazioni di origine Asiatica, all'interno della zona artica, rispetto a quelle di origine Caucasica ed Africana. Si pensa che i rapidi cambiamenti nelle regioni codificanti proteine dell'mtDNA, andando ad incidere sulla capacità dei mitocondri di produrre energia attraverso la fosforilazione ossidativa e generando calore, potrebbero essere stati selettivamente vantaggiosi nei flussi migratori di determinate popolazioni, che così si sono adattate ai nuovi climi ed ai nuovi regimi alimentari.

GENE	Lunghezza Tot (L.T.)	Polimorfismi osservati (P.O.)	Rapporto P.O./L.T.
CytB	1141bp	24	0.021
Subunità CO	3010bp	47	0.016
Subunità ATPsintasi	842bp	23	0.027
Subunità NADH	6145bp	111	0.018

Tab. IV.17 Variabilità nella regione codificante le subunità enzimatiche

4.6 Classificazione dei campioni analizzati in aplogruppi

L'insieme delle mutazioni riscontrate è stato utilizzato al fine di individuare gli aplotipi di appartenenza dei campioni, e, di conseguenza, di poterli classificare nei

rispettivi aplogruppi e sottoaplogruppi. La distribuzione così ottenuta è stata confrontata con quella della popolazione italiana e caucasica, per valutarne le implicazioni forensi.

Sulla base di tutte le informazioni così ricavate si è voluto inoltre verificare, a livello di altri studi, se i polimorfismi in questione fossero già stati individuati come SNPs utili per saggi forensi e/o se potessero concorrere alla messa a punto di tali pannelli, con riferimento specifico alla popolazione italiana.

Per ciascun campione si è scelto di suddividere i polimorfismi tra la regione di controllo e quella codificante (Tab. IV.20). Una prima analisi è stata effettuata considerando le informazioni provenienti dalla regione di controllo (Tab. IV.19); a tal fine ci si è avvalsi delle dettagliata descrizione dei cosiddetti *"Mutation Motifs"*, cioè delle stringhe caratteristiche del D-loop, classificate nel portale "mtDNAmanager" (http://mtmanager.yonsey.ac.kr) (Tab. IV.18). Grazie a queste, è stato possibile ripartire i campioni in 1 superaplogruppo (HV0), 1 aplogruppo (H) e vari sottoaplogruppi di H, U, J, T, N, W, X ed L (vedere tabella IV.21). Integrando l'informazione relativa alla regione di controllo con quella proveniente dal sequenziamento totale del genoma mitocondriale, è stata ottenuta, invece, una classificazione più dettagliata relativa ai vari sottoaplogruppi. I 40 campioni analizzati sono stati inseriti nella classificazione in aplogruppi comunemente accettata (Achilli et al, 2004, Finnila et al, 2001, Macaulay et al, 1999, Torroni et al. 1993, 1994, 1996, 1998). I marcatori distintivi degli aplogruppi sono quelli riportati nell'albero filogenetico complessivo presentato sul portale web http://www.phylotree.org (van Oven M e Kayser M. 2008) (fig. 4.20).

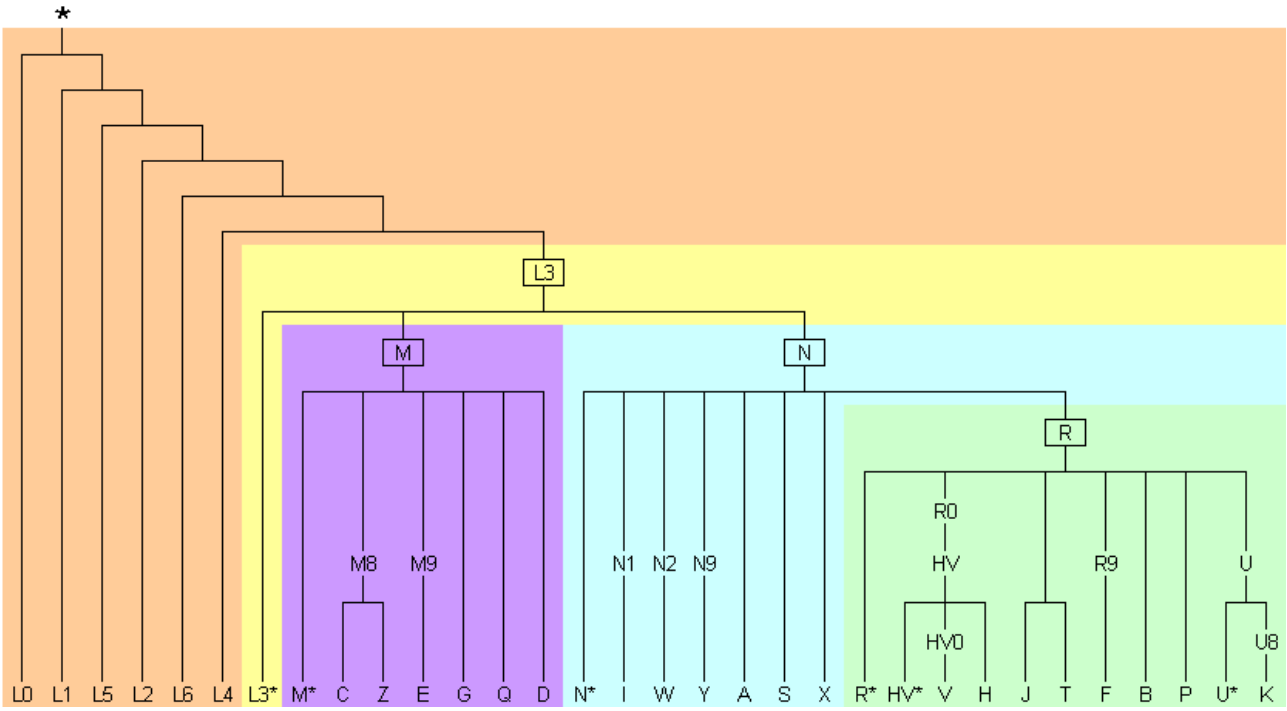


Fig. 4.20. Schema esemplificativo della filogenia mitocondriale, dettagliata nel PhiloTree di riferimento per il presente studio.

Per le informazioni relative all'assegnazione di ogni campione allo specifico aplogruppo/sottoaplogruppo, si è fatto riferimento alla bibliografia parimenti inserita nel portale. Le informazioni complessivamente ricavate sono riassunte nella tabella IV.21 ed illustrate graficamente nella figura 4.21.

Le eteroplasmie di lunghezza relative ai C-stretch, sia della regione HVI che della regione HVII, non sono state considerate ai fini dell'assegnazione ai diversi aplogruppi, dal momento che queste varianti non vengono, al momento, utilizzate ai fini dell'esclusione nei casi forensi (Holland eParsons, 1999; Stewart et al, 2001). Il PhiloTree, inoltre, non prende in considerazione la transversione 16182C e le possibili inserzioni di C nella posizione 16193. Esso non considera, inoltre, la mutazione T16519C: questa si trova in un tratto nella regione di controllo a tasso di variazione

HVI	**HVII**	**HVIII ecc.***	**Aplogruppo principale**	**Campioni che condividono l'aplotipo indicato**
16298	72		**HV0**	44, 96
	(73A)		**H**	65, 71, 76, 86, 100, 69, 104, 60, 21, 58, 66, 81
		16519 – 477	**H1c**	73
16235 - 16291			**H2a2b**	83
16311	152		**H2b**	98
16304		456	**H5**	82
16114			**H10a**	88
		55 – 57	**H15**	57
16256 - 16270		16399	**U5a1**	68, 62, 79
16256 - 16270C		16399	**U5a1**	28 (con retromutazione?)

HVI	**HVII**	**HVIII ecc.***	**Aplogruppo principale**	**Campioni che condividono l'aplotipo indicato**
16172 - 16219			**U6**	87
16343	150	16390	**U3a**	111, 105
(16356)	195 – 310	16519 – 499	**U4a2**	84
16179 – 16356	195	16519 – 499	**U4c1**	80
16189 – 16234			**U8b**	72
16069 – 16126	185 – 188 – 295	462 – 489	**J1c2**	67
16069 – 16126 - 16193	152 – 295	489	**J1d**	109
16126 – 16163 – 16186 – 16189 - 16294	(152 – 195)	16519	**T1a**	61
16126 – 16294 - 16296		16519	**T2**	64
16126 – 16294 – 16296 - 16304		16519	**T2b**	108, 107

HVI	HVII	HVIII ecc.*	Aplogruppo principale	Campioni che condividono l'aplotipo indicato
16145 – 16176R - 16223		16390 – 16519	**N1b**	70 (con possibile retromutazione della 16519)
16223 – 16292	189 – 194 – 195 – 204 – 207	16519	**W3'5**	85
16189 – 16278	195	16519	**X2**	59
16124 – 16223	152	523d-524d	**L3d**	63

Tab. IV.18 Mutation Motif di riferimento. Per ogni aplogruppo non si è tenuto conto dei polimorfismi nella posizione 73, 263, 315.1 e le altre inserzioni nel poliC di HVII; i polimorfismi sono tutti dati da transizioni, eccetto laddove specificatamente indicato. Le mutazioni tra parentesi sono state riscontrate nella maggior parte, ma non in tutte, le popolazioni appartenenti all'aplogruppo. * In questa colonna vengono indicati i polimorfismi in HVIII e tutti quelli non appartenenti né ad HVI, né ad HVII, ma comunque facenti parte della regione di controllo.

Campione	HG	Polimorfismi individuati nel D-loop								
GEA44	HV0	**T16298C**	**T72C**	T152C	*A263G*	*C309.1*	*C315.1*			
GEA96	HV0	G16129A	C16239A	**T16298C**	**T72C**	C194T	*A263G*	*C309.1*	*C315.1*	
GEA65	**H**	C16245T	A16309G	T16519C	**A73A**	T152C	*A263G*	*C309.1*	*C315.1*	
GEA71	**H**	T16093C	T16519C	**A73A**	*A263G*	*C315.1*				
GEA76	**H**	T16519C	**A73A**	T152C	*A263G*	*C309.1*	*C315.1*			
GEA86	**H**	T16368C	T16519C	**A73A**	*A263G*	TC310.2	*C315.1*			
GEA100	**H**	T16362C	T16519C	**A73A**	*A263G*	*C309.2*	*C315.1*			
GEA69	**H**	C16186T	T16189C	C16278T	T16519C	**A73A**	T152C	*A263G*	*C315.1*	T466C
GEA73	H1c	A16284G	C16355T	**T16519C**	*A263G*	*C309.1*	*C315.1*	**T477C**		
GEA83	H2a2b	**A16235G**	**C16291T**	A93G	*A263G*	*C309.1*	*C315.1*			
GEA98	H2b	T16209C	**T16311C**	**T152C**	*A263G*	*C309.1*	*C315.1*			
GEA82	H5	**T16304C**	T199C	*A263G*	*C309.2*	*C315.1*	**C456T**			
GEA88	H10a	**C16114T**	T16519C	*A263G*	*C309.2*	*C315.1*				
GEA104	**H**	C16188T	T16519C	**A73A**	T152C	T195C	A200G	*A263G*	*C309.1*	*C315.1*
GEA60	**H**	T16297C	**A73A**	T146C	*A263G*	*C309.1*	*C315.1*			

Campione	HG	Polimorfismi individuati nel D-loop										
GEA21	**H**	G16319A	T16519C	**A73A**	C182T	*A263G*	*C309.1*	*C315.1*				
GEA57	H15	**T55C**	**T57G**	*A263G*	*C309.1*	*C315.1*						
GEA58	**H**	**A73A**	*A263G*	*C315.1*								
GEA66	**H**	**A73A**	*A263G*	*C309.1*	*C315.1*							
GEA81	**H**	A16183C	T16189C	T16519C	**A73A**	*A263G*	*C315.1*					
GEA68	U5a1	**C16256T**	**C16270T**	**A16399G**	A73G	*A263G*	*C315.1*	T455C	T460C			
GEA28	U5a1	T16172C	C16221T	**C16256T**	C16270C	**A16399G**	A73G	*A263G*	*C309.1*	*C315.1*		
GEA62	U5a1	**C16256T**	**C16270T**	**A16399G**	A73G	*A263G*	*C309.1*	*C315.1*	**(573.1C) o**	A574C		
GEA79	U5a1	C16192T	**C16256T**	**C16270T**	C16294T	T16311C	**A16399G**	A73G	*A263G*	*C315.1*		
GEA87	U6	**T16172C**	**A16219G**	A73G	*A263G*	*C315.1*						
GEA111	U3a	**A16343G**	**G16390A**	T16519C	A73G	**C150T**	T152C	*A263G*	*C309.1*	*C315.1*		
GEA105	U3a	T16311C	**A16343G**	T16356C	**G16390A**	T16519C	A73G	**C150T**	*A263G*	*C309.1*	*C315.1*	A508G
GEA84	U4a2	**T16356C**	**T16519C**	A73G	T152C	**T195C**	*A263G*	**T310C**	**G499A**			
GEA80	U4c1	**C16179T**	**T16356C**	**T16519C**	A73G	T152C	**T195C**	*A263G*	*C309.1*	*C315.1*	**G499A**	
GEA72	U8b	T16172C	A16183C	**T16189C**	**C16234T**	T16311C	T16519C	A73G	C150T	T195C	*A263G*	*C315.1*

Campione	HG	Polimorfismi individuati nel D-loop										
GEA67	J1c2	**C16069T**	T16092C	**T16126C**	A73G	**G185A**	**A188G**	G228A	*A263G*	**C295T**	*C315.1*	**C462T**
		T489C										
GEA109	J1d	**C16069T**	**T16126C**	C16186T	**C16193T**	A16300G	A73G	**T152C**	A214G	*A263G*	**C295T**	*C309.1*
		C315.1	C462T	**T489C**								
GEA61	T1a	**T16126C**	**A16163G**	T16172C	**C16186T**	**T16189C**	**C16294T**	T16298C	A16399G	**T16519C**	A73G	A200G
		A263G	*C309.1*	*C315.1*								
GEA64	T2	**T16126C**	**C16294T**	**C16296T**	**T16519C**	A73G	*A263G*	*C315.1*				
GEA108	T2b	**T16126C**	C16270T	**C16294T**	**C16296T**	**T16304C**	**T16519C**	A73G	C151T	*A263G*	*C315.1*	
GEA107	T2b	**T16126C**	C16270T	**C16294T**	**C16296T**	**T16304C**	**T16519C**	A73G	C151T	*A263G*	*C315.1*	
GEA70	N1b	G16129A	**G16145A**	**C16176G**	**C16223T**	C16291T	**G16390A**	T16519T	A73G	T152C	G185A	A188G
		A263G	*C309.1*	*C315.1*								
GEA85	W3'5	**C16223T**	**C16292T**	C16320T	**T16519C**	A73G	**A189G**	**C194T**	**T195C**	**T204C**	**G207A**	*A263G*
		C309.1	*C315.1*									
GEA59	X2	**T16189C**	A16215G	C16223T	**C16278T**	C16294T	G16390A	**T16519C**	A73G	T152C	A153A	**T195C**
		A263G	*C309.1*	*C315.1*								
GEA63	L3d	**T16124C**	**C16223T**	T16519C	A73G	**T152C**	*A263G*	*C315.1*	**(523d-524d)**			

Tab. IV.19 Individuazione dei *mutation motifs* nel D-loop dei 40 campioni esaminati. In verde, i polimorfismi relativi ad HVI, in giallo quelli relativi ad HVII, in azzurro quelli relativi ad HVIII ed in fucsia le possibili retromutazioni (vedere testo). I polimorfismi non evidenziati fanno parte del D-loop ma si trovano al di fuori delle tre regioni ipervariabili; quelli in corsivo sono non utili per l'analisi (vedere testo). In grassetto sono stati indicati i polimorfismi caratteristici dei *mutation motifs* e/o gli aplogruppi presumibili, non individuabili in questa prima analisi.

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16298C**	**T72C**	**T152C**	*A263G*	*C309.1*	*C315.1*					
GEA44												
V	***Coding Region***	A750G	A1438G	A2706G	*N3107del*	**G4580A**	A4769G	C7028T	**A8050G**	*A8860G*	**G9380A**	**G11719G**
		C14766C	*A15326G*	**C15904T**								
	Control Region	**G16129A**	**C16239A**	**T16298C**	**T72C**	**C194T**	*A263G*	*C309.1*	*C315.1*			
GEA96												
V	***Coding Region***	A750G	A1438G	A2706G	*N3107del*	**G4580A**	A4769G	C7028T	*A8860G*	**G11719G**	**C12813T**	**C14766C**
		A15326G	**C15455T**	**C15904T**								
	Control Region	**C16245T**	**A16309G**	T16519C	**A73A**	**T152C**	*A263G*	*C309.1*	*C315.1*			
GEA65												
H	***Coding Region***	A750G	A1438G	**A2706A**	*N3107del*	A4769G	**C7028C**	**T7621C**	**T8639C**	*A8860G*	*A15326G*	
	Control Region	**T16093C**	T16519C	**A73A**	*A263G*	*C315.1*						
GEA71												
H	***Coding Region***	A750G	A1438G	**T2159C**	**A2706A**	*N3107del*	**G3918A**	A4769G	**A5951G**	**C7028C**	**T7094C**	**A8701G**
		A8860G	**A14193G**	*A15326G*								
	Control Region	T16519C	**A73A**	**T152C**	*A263G*	*C309.1*	*C315.1*					
GEA76												
H	***Coding Region***	A750G	A1438G	**A2706A**	*N3107del*	A4769G	**A6647G**	**C7028C**	*A8860G*	**G9196A**	**C13785T**	*A15326G*
		G15927A										
	Control Region	**T16368C**	T16519C	**A73A**	*A263G*	*TC310.2*	*C315.1*					
GEA86												
H	***Coding Region***	A750G	A1438G	**A2706A**	*N3107del*	**C3435T**	A4769G	**C7028C**	*A8860G*	*A15326G*		

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16362C**	T16519C	**A73A**	*A263G*	*C309.2*	*C315.1*					
GEA100												
H1	***Coding Region***	A750G	A1438G	**G3010A**	*N3107del*	A4769G	*A8860G*	*A15326G*				
	Control Region	**C16186T**	**T16189C**	**C16278T**	T16519C	**A73A**	**T152C**	*A263G*	*C315.1*	**T466C**		
GEA69												
H1a-b'f'g'k'q	***Coding Region***	A750G	A1438G	**G3010A**	*N3107del*	A4769G	*A8860G*	**G10427A**	*A15326G*			
	Control Region	**A16284G**	**C16355T**	T16519C	*A263G*	*C309.1*	*C315.1*	**T477C**				
GEA73												
H1c	***Coding Region***	A750G	A1438G	**G3010A**	*N3107del*	A4769G	**G5744A**	*A8860G*	**G13759A**	*A15326G*		
	Control Region	**A16235G**	**C16291T**	**A93G**	*A263G*	*C309.1*	*C315.1*					
GEA83												
H2a2b1	***Coding Region***	**A1438A**	*N3107del*	**A4769A**	*A8860G*	**A14037G**	*A15326G*					
	Control Region	**T16209C**	**T16311C**	**T152C**	*A263G*	*C309.1*	*C315.1*					
GEA98												
H2b	***Coding Region***	A750G	**A1438A**	*N3107del*	A4769G	**T8598C**	*A8860G*	*A15326G*				
	Control Region	**T16304C**	**T199C**	*A263G*	*C309.2*	*C315.1*	**C456T**					
GEA82												
H5a1	***Coding Region***	A750G	A1438G	*N3107del*	**C3804T**	**T4336C**	A4769G	**A6800G***	*A8860G*	*A15326G*	**T15511C**	**C15833T**

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16114T**	T16519C	*A263G*	*C309.2*	*C315.1*						
GEA88												
H10a1	***Coding Region***	A750G	A1438G	*N3107del*	**T3398C**	**T4216C**	A4769G	**T8419C**	**A8531G**	A8860G	**G13711A**	**T14470A**
		A14548G	*A15326G*									
	Control Region	**A16037G**	**C16188T**	T16519C	**A73A**	**T152C**	**T195C**	**A200G**	*A263G*	*C309.1*	*C315.1*	
GEA104												
H11'12	***Coding Region***	**G709A**	A750G	A1438G	*N3107del*	A4769G	**T7270C**	*A8860G*	**G11914A**	*A15326G*		
	Control Region	**T16297C**	**A73A**	**T146C**	*A263G*	*C309.1*	*C315.1*					
GEA60												
H13a1a1a	***Coding Region***	**C679A**	A750G	A1438G	**C2259T**	*N3107del*	**A4745G**	A4769G	**G7337A**	*A8860G*	**T13326C**	**C13680T**
		C14872T	*A15326G*									
	Control Region	**G16319A**	T16519C	**A73A**	**C182T**	*A263G*	*C309.1*	*C315.1*				
GEA21												
H13a2a	***Coding Region***	**G709A**	A750G	**A1008G**	**A1438A**	**C2259T**	*N3107del*	A4769G	**G8152A**	*A8860G*	**T8945C**	**G9575A**
Allele raro		**A12662G**	**A13105G**	**C14872T**	*A15326G*							
	Control Region	**T55C**	**T57G**	*A263G*	*C309.1*	*C315.1*						
GEA57												
H15b	***Coding Region***	A750G	**A1117G**	A1438G	*N3107del*	**T3847C**	A4769G	**T6253C**	*A8860G*	*A15326G*		
	Control Region	**A73A**	*A263G*	*C315.1*								
GEA58												
H33	***Coding Region***	A750G	A1438G	*N3107del*	A4769G	*A8860G*	**A9115G**	**G11447A**	*A15326G*			

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**A73A**	*A263G*	*C309.1*	*C315.1*							
GEA66												
H33	***Coding Region***	A750G	A1438G	*N3107del*	A4769G	*A8860G*	**A9115G**	**G11447A**	*A15326G*			
	Control Region	A16183C	**T16189C**	T16519C	**A73A**	*A263G*	*C315.1*					
GEA81												
H35	***Coding Region***	A750G	A1438G	*N3107del*	**C3342T**	**A4227G**	A4769G	**T7624C**	**T8047C**	*A8860G*	**G12618A**	*A15326G*
	Control Region	**C16192C**	**C16256T**	**C16270T**	**A16399G**	**A73G**	*A263G*	*C315.1*	**T455C**	**T460C**		
GEA68												
	Coding Region	A750G	A1438G	**T1700C**	A2706G	*N3107del*	**T3197C**	A4769G	**T5495C**	C7028T	**A8296G**	*A8860G*
U5a1a1		**G9477A**	**A10978G**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**	*A15326G*
		A15924G										
	Control Region	**T16172C**	**C16192C**	**C16221T**	**C16256T**	**C16270C**	**A16399G**	**A73G**	*A263G*	*C309.1*	*C315.1*	
GEA28												
U5a1a1	***Coding Region***	**T485C**	A750G	A1438G	**T1700C**	A2706G	*N3107del*	**T3197C**	A4769G	**T5495C**	C7028T	**T8705C**
Retromutazione		*A8860G*	**G9477A**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**	*A15326G*
		A15924G										
	Control Region	**C16192C**	**C16256T**	**C16270T**	**A16399G**	**A73G**	*A263G*	*C309.1*	*C315.1*	**573.1C**		
GEA62												
U5a1a2a	***Coding Region***	A750G	A1438G	**T1700C**	A2706G	*N3107del*	**T3197C**	A4769G	**A5319G**	**A6629G**	**T6719C**	C7028T
		A8860G	**G9477A**	**T10810C**	**A11467G**	**G11719A**	**A12308G**	**C12346T**	**G12372A**	**A13470G**	**T13617C**	**C14766T**
		A14793G	**A15218G**	*A15326G*								

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16192T**	**C16256T**	**C16270T**	**C16294T**	**T16311C**	**A16399G**	**A73G**	*A263G*	*C315.1*		
GEA79												
U5a1f	***Coding Region***	A750G	A1438G	A2706G	**C2911T**	*N3107del*	**T3197C**	**A4164G**	A4769G	**G5585A**	**G6023A**	C7028T
		A7569G	*A8860G*	**G9477A**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**
		A15326G										
	Control Region	**T16172C**	**A16219G**	**C16278C**	**A73G**	*A263G*	*C315.1*					
GEA87												
U6a7a1	***Coding Region***	A750G	**T794A**	**T1193C**	A1438G	**A1692T**	A2706G	*N3107del*	**A3348G**	A4769G	**A5120G**	**G5471A**
Hot spot		**T6512C**	C7028T	**G7805A**	**T8473C**	*A8860G*	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**A14179G**	**C14766T**
		G15043A	*A15326G*	**T15530C**	**C15632T**							
	Control Region	**A16343G**	**G16390A**	T16519C	**A73G**	**C150T**	**T152C**	*A263G*	*C309.1*	*C315.1*		
GEA111												
U3a	***Coding Region***	A750G	A1438G	**A1811G**	**A2294G**	A2706G	*N3107del*	**T4452C**	**T4703C**	A4769G	**C5424T**	**T6152C**
		T6297C	**C6518T**	C7028T	*A8860G*	**G9266A**	**A10506G**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**A12930G**
		C13934T	**A14139G**	**C14766T**	*A15326G*	**T15454C**						
	Control Region	**T16311C**	**A16343G**	**T16356C**	**G16390A**	T16519C	**A73G**	**C150T**	*A263G*	*C309.1*	*C315.1*	**A508G**
GEA105												
U3a1	***Coding Region***	A750G	A1438G	**A1811G**	**A2294G**	A2706G	**G3010A**	*N3107del*	**T4703C**	A4769G	**C6518T**	C7028T
		A8860G	**G9266A**	**A10506G**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**C13934T**	**A14139G**	**C14766T**	*A15326G*
		T15454C	**G15734A**									
	Control Region	**T16356C**	T16519C	**A73G**	**T152C**	**T195C**	*A263G*	**T310C**	**G499A**			
GEA84												
U4a2	***Coding Region***	A750G	A1438G	**A1811G**	A2706G	*N3107del*	**T4646C**	A4769G	**T5999C**	**A6047G**	C7028T	**C8818T**
		A8860G	**T10493C**	**C11332T**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**C14620T**	**C14766T**	*A15326G*	**T15693C**

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16179T**	**T16356C**	T16519C	**A73G**	**T152C**	**T195C**	*A263G*	*C309.1*	*C315.1*	**G499A**	
GEA80												
U4c1a	***Coding Region***	**G709A**	A750G	**A827G**	A1438G	**A1811G**	A2706G	*N3107del*	**T4646C**	A4769G	**A4811G**	**T5999C**
		A6047G	**A6146G**	C7028T	*A8860G*	**T9070G**	**T10907C**	**T11009C**	**C11332T**	**A11467G**	**G11719A**	**A12308G**
		G12372A	**C14620T**	**C14766T**	**C14866T**	*A15326G*	**T15693C**					
	Control Region	**T16172C**	A16183C	**T16189C**	**C16234T**	**T16311C**	T16519C	**A73G**	**C150T**	**T195C**	*A263G*	*C315.1*
GEA72												
U8b1	***Coding Region***	A750G	A1438G	**A1811G**	**A1943G**	A2706G	*N3107del*	**A3480G**	**T3540C**	**A4021T**	A4769G	**C6546T**
		A6599G	C7028T	*A8860G*	**G9055A**	**T9698C**	**A10086G**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**G12771A**
		A13269G	**C14167T**	**C14766T**	**G14831A**	*A15326G*						
	Control Region	**C16069T**	**T16092C**	**T16126C**	**A73G**	**G185A**	**A188G**	**G228A**	*A263G*	**C295T**	*C315.1*	**C462T**
GEA67		**T489C**										
J1c2												
	Coding Region	A750G	A1438G	A2706G	**G3010A**	*N3107del*	**T4216C**	**G4655A**	A4769G	**A5198G**	**G5231A**	C7028T
		A8860G	**A10398G**	**A11251G**	**G11719A**	**C12346T**	**A12612G**	**A12729G**	**G13708A**	**C14766T**	**T14798C**	*A15326G*
		C15452A										
	Control Region	**C16069T**	**T16126C**	**C16186T**	**C16193T**	**A16300G**	**A73G**	**T152C**	**A214G**	*A263G*	**C295T**	*C309.1*
GEA109		*C315.1*	**C462T**	**T489C**								
J1d1												
	Coding Region	A750G	**G1007A**	A1438G	**A1636G**	A2706G	**G3010A**	*N3107del*	**T4216C**	A4769G	C7028T	**G7789A**
		A7963G	*A8860G*	**A9336G**	**A10398G**	**A11251G**	**G11719A**	**G12528A**	**A12612G**	**G13135A**	**G13708A**	**C14766T**
		G15043A	*A15326G*	**C15452A**								

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16126C**	**A16163G**	**T16172C**	**C16186T**	**T16189C**	**C16294T**	**T16298C**	**A16399G**	T16519C	**A73G**	**A200G**
GEA61		*A263G*	*C309.1*	*C315.1*								
T1a												
	Coding Region	**G709A**	A750G	A1438G	**G1888A**	A2706G	*N3107del*	**T4216C**	A4769G	**A4917G**	**C5345T**	**T5806C**
		C7028T	**G8697A**	*A8860G*	**T10463C**	**A11251G**	**C11356T**	**G11719A**	**C12633A**	**G13368A**	**C14766T**	**G14905A**
		A15326G	**C15452A**	**A15607G**	**G15928A**							
	Control Region	**T16126C**	**C16294T**	**C16296T**	T16519C	**A73G**	*A263G*	*C315.1*				
GEA64												
T2												
	Coding Region	**G709A**	A750G	A1438G	**G1888A**	**T2010C**	A2706G	*N3107del*	**G4113A**	**T4216C**	A4769G	**A4917G**
		G5471A	**C6445T**	C7028T	**G8697A**	*A8860G*	**T9615C**	**T10463C**	**A11251G**	**G11719A**	**A11812G**	**T12408C**
		T13260C	**G13368A**	**A14233G**	**G14323A**	**A14605G**	**C14766T**	**G14905A**	*A15326G*	**C15452A**	**A15607G**	**G15928A**
	Control Region	**T16126C**	**C16270T**	**C16294T**	**C16296T**	**T16304C**	T16519C	**A73G**	**C151T**	*A263G*	*C315.1*	
GEA108												
T2b3	***Coding Region***	**G709A**	A750G	**G930A**	A1438G	**G1888A**	A2706G	*N3107del*	**T4216C**	A4769G	**C4843T**	**A4917G**
		G5147A	C7028T	**G8697A**	*A8860G*	**A9007G**	**T10463C**	**A10750G**	**G10993A**	**A11251G**	**G11719A**	**A11812G**
		G13368A	**A14233G**	**C14766T**	**G14905A**	*A15326G*	**C15452A**	**A15607G**	**G15928A**			
	Control Region	**T16126C**	**C16270T**	**C16294T**	**C16296T**	**T16304C**	T16519C	**A73G**	**C151T**	*A263G*	*C315.1*	
GEA107												
T2b3	***Coding Region***	**G709A**	A750G	**G930A**	A1438G	**G1888A**	A2706G	*N3107del*	**T4216C**	A4769G	**C4843T**	**A4917G**
		G5147A	C7028T	**G8697A**	*A8860G*	**A9007G**	**T10463C**	**A10750G**	**G10993A**	**A11251G**	**G11719A**	**A11812G**
		G13368A	**A14233G**	**C14766T**	**G14905A**	*A15326G*	**C15452A**	**A15607G**	**G15928A**			

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**G16129A**	**G16145A**	**C16176G**	**C16223T**	**C16291T**	**G16390A**	**A73G**	**T152C**	**G185A**	**A188G**	*A263G*
GEA70		*C309.1*	*C315.1*									
N1b1d												
	Coding Region	A750G	A1438G	**G1598A**	**C1703T**	**G1719A**	**C2639T**	A2706G	*N3107del*	**C3921A**	A4769G	**C4960T**
		G5471A	**T5528C**	C7028T	**G8251A**	**A8261G**	**C8410T**	**C8472T**	**A8701A**	**T8763C**	**A8836G**	*A8860G*
		C9335T	**T9540T**	**T10238C**	**A10398A**	**T10873T**	**A11362G**	**G11719A**	**G12501A**	**C12705T**	**A12822G**	**A13866G**
		C14766T	**G15301G**	*A15326G*								
	Control Region	**C16223T**	**C16292T**	**C16320T**	T16519C	**A73G**	**A189G**	**C194T**	**T195C**	**T204C**	**G207A**	*A263G*
GEA85		*C309.1*	*C315.1*									
W3'6												
	Coding Region	**G709A**	A750G	**T1243C**	A1438G	A2706G	*N3107del*	**A3505G**	A4769G	**G5046A**	**G5460A**	C7028T
		G8251A	*A8860G*	**G8994A**	**C11674T**	**G11719A**	**A11947G**	**T12414C**	**C12705T**	**C14766T**	*A15326G*	**G15884C**
	Control Region	**T16189C**	**A16215G**	**C16223T**	**C16278T**	**C16294T**	**G16390A**	T16519C	**A73G**	**T152C**	**A153A**	**T195C**
		A263G	*C309.1*	*C315.1*								
GEA59												
X2		A750G	A1438G	**G1719A**	A2706G	*N3107del*	**T4561C**	A4769G	**T6221C**	**C6371T**	**A6791G**	C7028T
Hot spot	***Coding Region***	**T8503C**	*A8860G*	**T10248C**	**G11719A**	**C12403T**	**C12705T**	**G13708A**	**T13899C**	**A13966G**	**T14470C**	**A14755G**
		C14766T	*A15326G*									

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16124C**	**C16223T**	T16519C	**A73G**	**T152C**	*A263G*	*C315.1*	**(523d-524d)**			
GEA63												
L3d1b1	***Coding Region***	A750G	**T921C**	A1438G	A2706G	*N3107del*	A4769G	**G5046A**	**G5147A**	**A6272G**	**T6680C**	C7028T
		A7424G	**T8618C**	**A8701G**	*A8860G*	**T9540C**	**A10398G**	**T10873C**	**G11719A**	**C12705T**	**C12745T**	**A13105G**
		T13886C	**C14284T**	**C14766T**	**G15301A**	*A15326G*						

*eteroplasmia

Tab. IV.20 Individuazione degli SNPs nel D-loop e nella regione codificante dei 40 campioni esaminati. In verde, i polimorfismi del D-loop, già individuati come inclusi nei *mutation motifs*; in azzurro gli ulteriori polimorfismi, inclusi nell'albero filogenetico ed utilizzati per l'assegnazione definitiva ai vari aplogruppi (indicati sotto ogni campione); in fucsia le possibili retromutazioni (vedere testo); in grigio i polimorfismi la cui assenza determina l'assegnazione ad un dato aplogruppo: a titolo esemplificativo ne vengono indicati solo alcuni, sottintendendone la presenza nei relativi sottoaplogruppi e/o aplogruppi correlati: per ulteriori dettagli si rinvia al testo; in carattere *corsivo* i polimorfismi non considerati nell'albero filogenetico ed, in generale, nell'assegnazione ai vari aplogruppi; in carattere normale i polimorfismi non discriminanti o già abbondantemente descritti (vedere anche tab…); in **rosso** i polimorfismi presenti nel campione in esame come discriminanti per un dato aplogruppo in alcuni casi, e come mutazioni private in altri; in **blu** i polimorfismi rappresentanti mutazioni private nei relativi soggetti.

relativamente basso, ma possiede il tasso di sostituzione evolutivo più alto in tutto il genoma mitocondriale e quindi è molto utile per l'ulteriore discriminazione a livello di regione di controllo. Si è scelto, infine, di non considerare le transizioni A263G, A8860G e A15326G condivise da tutti i campioni.

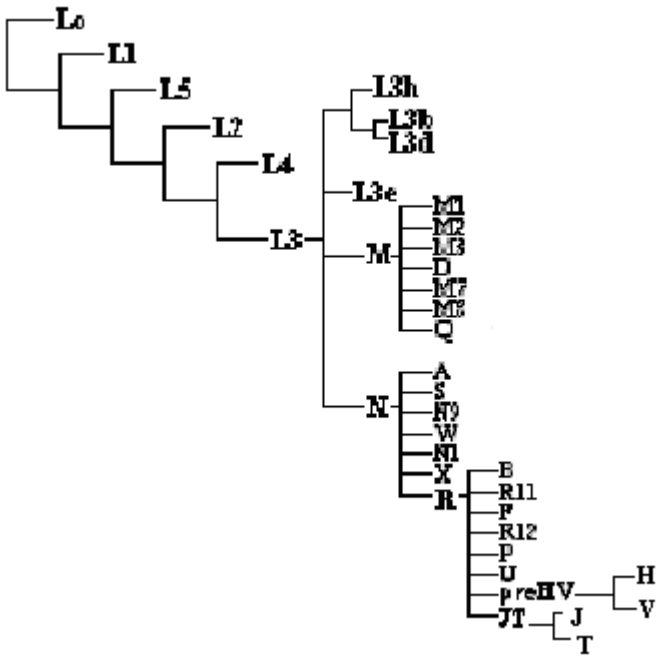


Fig. 4.21 Albero raffigurante gli aplogruppi principali individuati.

La tabella IV.21 mostra i risultati della classificazione completa, sia quella ottenuta mediante la sola informazione proveniente dalla regione di controllo, sia quella ottenuta sequenziando anche la regione codificante.

Campione	Hg (D-loop)	Hg (full mtDNA)	N°osservazioni	Totali
44	**HVO**	**V**	**1**	**2**
96	**HVO**	**V**	**1**	
65,71,76,86	**H**	**H**	**4**	**18**
100	H	H1	**1**	
69	H	H1a-b'f'g'k'q*	**1**	
73	H1c	H1c	**1**	
83	H2a2ab	H2a2b1	**1**	
98	H2b	H2b	**1**	
82	H5	H5a1	**1**	
88	H10a	H10a1	**1**	
104	H	H11'12*	**1**	
60	H	H13a1a1a	**1**	
21	H	H13a2a	**1**	
57	H15	H15b	**1**	
58,66	H	H33	**2**	
81	H	H35	**1**	
68	**U5a1**	**U5a1a1**	**1**	**10**
28	**U5a1**	**U5a1a1**	**1**	
62	**U5a1**	**U5a1a2a**	**1**	
79	**U5a1**	**U5a1f**	**1**	
87	**U6**	**U6a7a1**	**1**	
111	**U3a**	**U3a**	**1**	
105	**U3a**	**U3a1**	**1**	
84	**U4a2**	**U4a2**	**1**	
80	**U4c1**	**U4c1a**	**1**	
72	**U8b**	**U8b1**	**1**	
67	**J1c2**	**J1c2**	**1**	**2**
109	**J1d**	**J1d1**	**1**	
61	**T1a**	**T1a**	**1**	**4**
64	**T2**	**T2**	**1**	
108	**T2b**	**T2b3**	**1**	
107	**T2b**	**T2b3**	**1**	
70	**N1b**	**N1b1d**	**1**	***1***
85	**W3'5**	**W3'6***	**1**	**1**
59	**X2**	**X2**	**1**	**1**
63	**L3d**	**L3d1b1**	**1**	***1***

Tab. IV.21 Classificazione in aplogruppi dei 40 campioni analizzati. In grassetto sono evidenziati gli aplogruppi principali (V, H) ed i sottoaplogruppi dei principali aplogruppi (U, J, T, N, W, X, L3) non rappresentati nella popolazione in esame; in carattere semplice i sottoaplogruppi dell'aplogruppo H qui rappresentato. Sono evidenziati in colore fucsia quei campioni la cui assegnazione ad un dato aplogruppo ha richiesto la valutazione di condizioni particolari (vedere testo) ed in verde quelli risolti dal sequenziamento del DNA codificante. *Sottoaplogruppo intermedio, ancora in via di classificazione sul PhiloTree (vedi testo).

I campioni per i quali non viene specificato il sottoaplogruppo, ma che vengono inseriti negli aplogruppi principali non sono ascrivibili a nessuna delle sottoclassi stabilite in letteratura. Tra questi, ve ne sono alcuni che potrebbero essere definiti "intermedi", e precisamente: a) GEA69 (fig. 4.22), tipicamente H1, inserito sull'albero filogenetico tra H1a ed H1c, per via del polimorfismo in 16189, che lo colloca come base per il cluster H1b, H1f, H1g H1k e H1q.

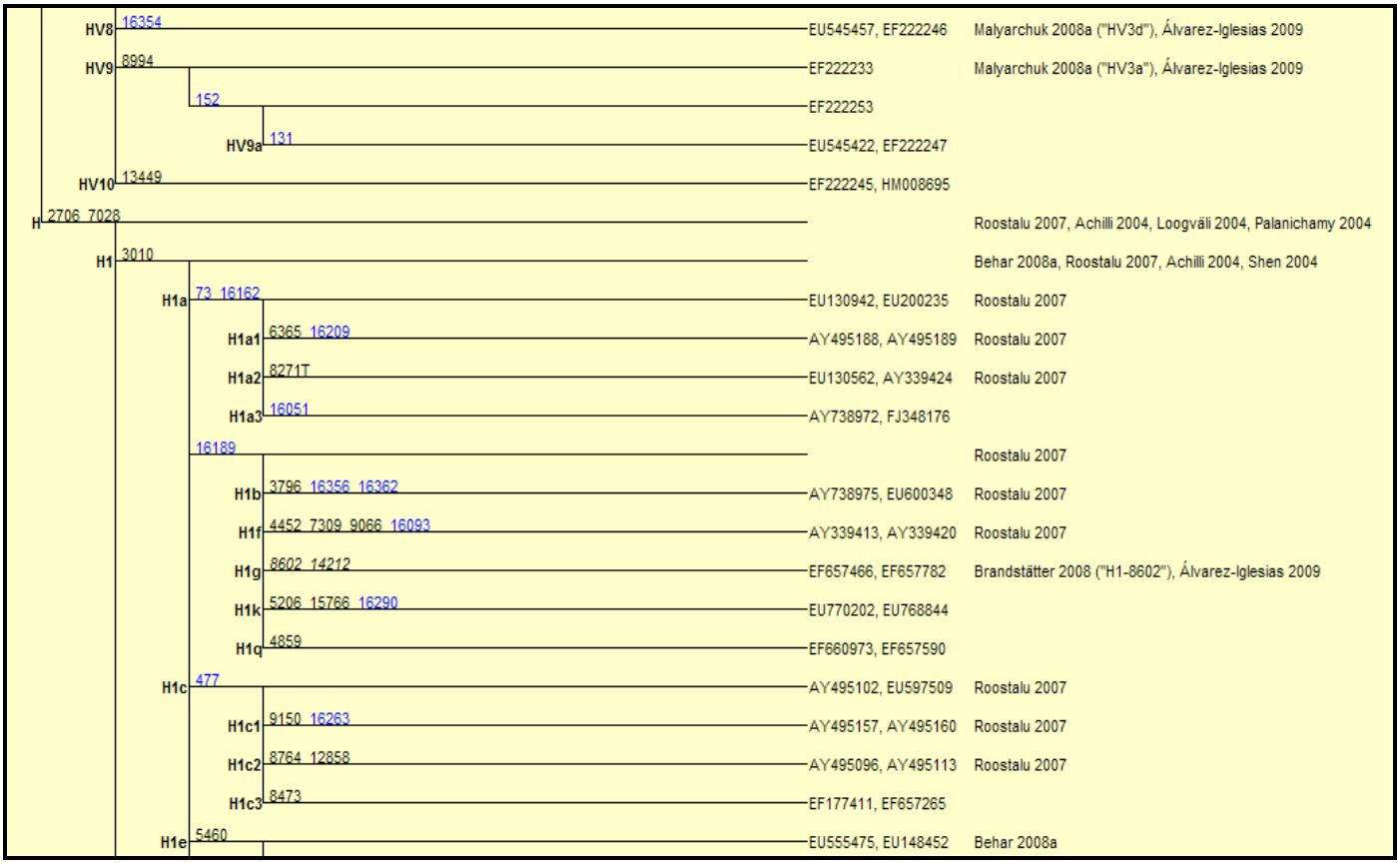


Fig. 4.22 Posizione ambigua di GEA69: un pre_sottoaplogruppo del cluster H1b, H1f, H1g, H1k, H1q. Vengono indicate in azzurro le posizioni relative alla regione di controllo, ed in nero quelle relative alla regione codificante.

b) GEA104 (fig. 4.23), inserito sull'albero filogenetico tra H10 ed H13, per effetto del polimorfismo in posizione 195, ma ancora in via di definizione.

Fig. 4.23 Posizione ambigua di GEA104: un pre_sottoaplogruppo del cluster H11/H12. Vedere anche legenda fig. 4.22.

c) GEA85 (fig. 4.24), chiaramente appartenente all'aplogruppo W, è comunque inserito sull'albero filogenetico in cima al cluster W3,W4,W5 e W6, per effetto del polimorfismo in posizione 194, ma ancora in via di definizione.

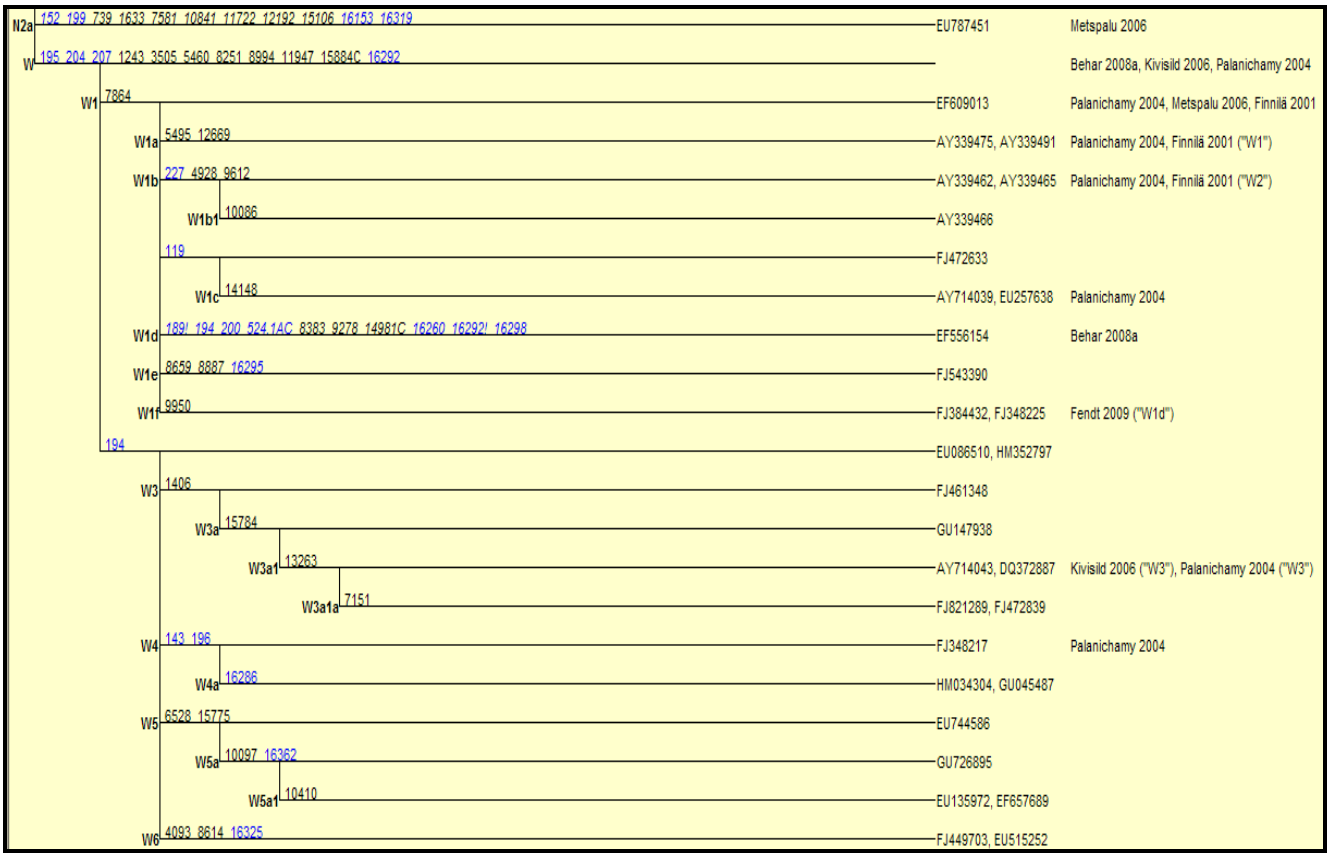


Fig. 4.24 Posizione ambigua di GEA85: un pre_sottoaplogruppo del cluster W3/W4/W5/W6. Vedere anche legenda fig. 4.22.

Gli aplogruppi Caucasici H e V sono aplogruppi “fratelli”, riscontrati ad una frequenza del 45% circa e del 5% rispettivamente, all’interno delle popolazioni caucasiche europee e statunitensi (Torroni et al, 1996; Macaulay et al, 1999a; Allard et al, 2002). Il dato del presente studio è in perfetta linea con tale riscontro. Il cluster HV è raggruppato dai polimorfismi diagnostici della regione codificante 11719G e 14766C (Macaulay et al, 1999a; Saillard et al, 2000; Achilli et al, 2004; Palanichamy et al, 2004; Torroni et al, 2006; Kivisild et al, 2006; Behar et al, 2008a; Brandstätter et al, 2008; Álvarez-Iglesias et al, 2009). L’aplogruppo H viene definito dalle mutazioni della regione codificante nelle posizioni 2706 e 7028; generalmente vi è anche la presenza del polimorfismo della regione di controllo 73A, con contemporanea mancanza di

polimorfismi specifici, associati ad altri aplogruppi (Torroni et al, 1996). Tuttavia sono state descritte in letteratura un numero di reversioni nella posizione 73 (Torroni et al, 1996; Macaulay et al, 1999a; Allard et al, 2002; Roostalu et al, 2007; Achilli et al, 2004; Loogväli et al, 2004; Palanichamy et al, 2004). Oltre alle mutazioni specifiche del superaplogruppo HV, l'aplogruppo V è caratterizzato da mutazioni specifiche nella regione codificante, nelle posizioni 4580 (G → A) e 15904 (C → T) e mutazioni specifiche nella regione di controllo in 72 (T → C) e 16298 (T → C). Quest'ultima posizione, peraltro un sito relativamente "veloce" ("fast") a livello mutazionale (Wakeley 1993; Hasegawa et al, 1993; Torroni et al, 2001; Allard et al, 2002), è comunque l'unica, nell'ambito dell'analisi strettamente limitata ad HVI/HVII che distingue gli aplogruppi V dagli H (la posizione 72, specifica dell'aplogruppo V, è al di fuori del range di HVII), pertanto l'omoplasia in tale posizione potrebbe dare un'inaccurata attribuzione ad un dato aplogruppo, affidandosi unicamente alle informazioni relative ad HVI ed HVII. Grazie al sequenziamento del tratto codificante è stato quindi possibile risolvere i due campioni HVO come appartenenti all'aplogruppo V. Per quanto riguarda l'aplogruppo H, la prima analisi ha individuato 18 campioni, di cui 12 non ulteriormente classificabili e 6 appartenenti a dei sottoaplogruppi; il completamento delle sequenze ha lasciato non ulteriormente classificati solo 4 dei 12 campioni, mentre i rimanenti 8 sono stati ulteriormente definiti (sebbene tra questi ve ne siano due "intermedi", come poc'anzi detto). Tra i 6 sottoaplogruppi individuati dalla prima analisi, 2 son rimasti tal quali (H1c ed H2b) mentre gli altri 4 sono stati meglio dettagliati dall'analisi del DNA codificante (vedere tabella). Nonostante l'esiguità del campione, la distribuzione dei sottoaplogruppi

H riflette quella delle popolazioni Caucasiche: il più rappresentato è risultato essere l'H1, con 3 hits, che è quello più frequente nelle popolazioni europee (Grignani et al, 2006; Roostalu et al, 2007), seguito dall'H2, con 2 hits, di origine Eurasiatica (Loogväli et al, 2004) e via via dagli altri. Un'osservazione particolare va fatta in merito al sottoaplogruppo H13a (2 hits in questo studio), frequente nel Daghestan, in Georgia ed in Europa (Roostalu et al, 2007): la stringa del D-loop, relativa alla sua definizione, non è stata riscontrata in alcun campione di questo studio: pur tuttavia l'analisi della regione codificante ha fatto ascrivere a tale aplogruppo ben due campioni. GEA60, in particolare, è del tutto simile alla sequenza AY339410 (Finnilä et al, 2001), a suo tempo classificata come H e del tutto priva del *mutation motif* 161218-16278-93 (http://mtmanager.yonsey.ac.kr); allo stesso sottoaplogruppo di GEA60 (H13a1a1a) è stata ascritta la sequenza AY738995 (Achilli et al, 2004), nella quale è presente solo la transizione in posizione 93. Un caso simile e doppiamente interessante riguarda il campione GEA21: esso presenta tutti i marcatori del sottoaplogruppo **H13a2a**, ma manca del polimorfismo A1438G. (fig. 4.25) Nel soggetto è presente l'allele raro A1438A, comunque descritto nei Caucasici (Coble et al, 2004); inoltre GEA21 ha riscontro in Gene Bank relativamente alla sequenza depositata EF556150 (Behar et al, 2008a), classificata come H13a2a1, ed alla sequenza AY713980 (Palanichamy et al, 2004), classificata come H; tutte e tre le sequenze non presentano il *mutaton motif* 16148-16256-16319-189-193-249. Ulteriori studi dovranno necessariamente chiarire se questo particolare sottoaplogruppo è effettivamente individuabile a mezzo del D-loop, e con

quali polimorfismidiscriminanti.

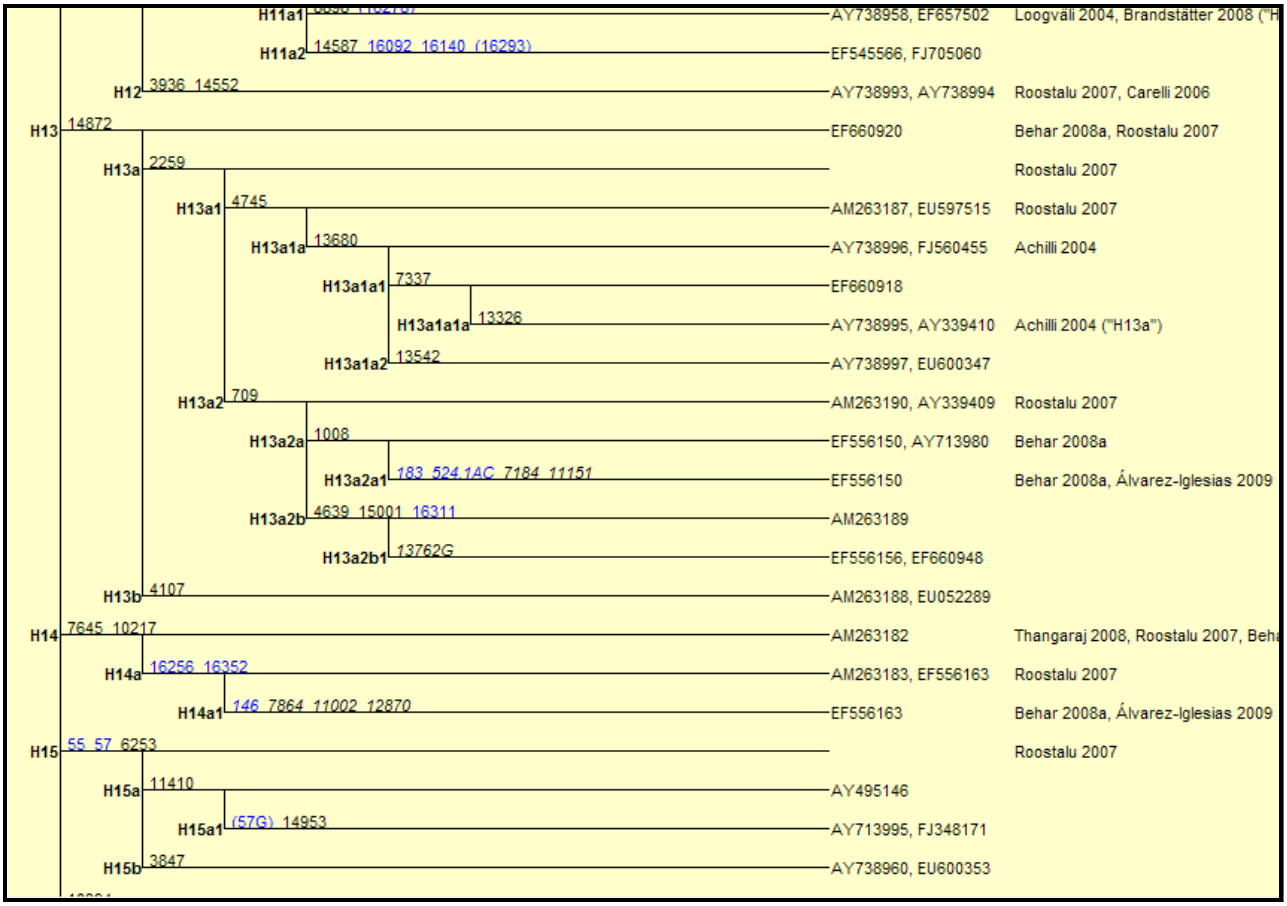


Fig. 4.25 Polimorfismi specifici per il sottoaplogruppo H13a2a

Infine, l'aplogruppo H33 è stato individuato in 2 soggetti, (GEA58 e GEA66), entrambi del Sud Italia, identici al 100% nel loro genoma mitocondriale: ciò lascia supporre che i medesimi abbiano un comune ascendente in linea materna (vedere Introduzione). Tutti gli altri aplogruppi e sottoaplogruppi sono rappresentati singolarmente.

L'aplogruppo caucasico K è un sottogruppo del cluster dell'aplogruppo U (Richards et al, 1998; Finnila et al, 2001) ed ha una frequenza di circa il 9% nei Nord Americani e negli Europei (Torroni et al, 1996; Allard et al, 2002). Esso viene definito dai polimorfismi della regione codificante: A1811G, G9055A, A12308G e G12372A, e

dai polimorfismi della non codificante: T16224C, T16311C e 73G (Torroni et al, 1996; Macaulay et al, 1999a; Finnila et al, 2001; Allard et al, 2002). Nel presente studio non sono stati individuati soggetti appartenenti a tale aplogruppo mentre, al contrario, sembra esservi un eccesso di soggetti con aplogruppo U rispetto a quanto osservato nei Caucasici (vedere oltre, tab. IV.22); in riferimento alle percentuali presenti nei Caucasici, le osservazioni del presente studio, riferite globalmente al superaplogruppo UK, non si discostano in maniera significativa da quanto riscontrabile nella popolazione Caucasica. L'analisi del D-loop ha individuato complessivamente 10 soggetti ascrivibili all'aplogruppo U; l'aplogruppo U5 è quello propriamente caucasico ed è definito dai polimorfismi in 3197C e 16270T. 3 soggetti appartengono al sottoaplogruppo U5a1 ed un quarto, GEA28, ascrivibile al medesimo sottoaplogruppo postulando una retromutazione nella posizione 16270. L'ulteriore definizione del campione ha in effetti dimostrato che questo sarebbe ascrivibile al sottoaplogruppo **U5a1a1** (fig. 4.26), ma manca del polimorfismo C16270T. Un altro campione, GEA68, appartiene al medesimo sottoplogruppo, parimenti definito mediante il sequenziamento totale, e possiede, invece, il polimorfismo in questione. In GEA28 è plausibile che si sia verificata una retromutazione, che ha riportato l'allele allo stato *wild*; tale fenomeno è stato curiosamente descritto nella popolazione Finlandese (Finnilä et al, 2000). Più in generale vale la pena ricordare che le retromutazioni consentono l'identificazione di sottoaplogruppi significativi in determinate aree, che possono rappresentare dei marcatori dei vari patterns di migrazione (Morelli et al, 2000).

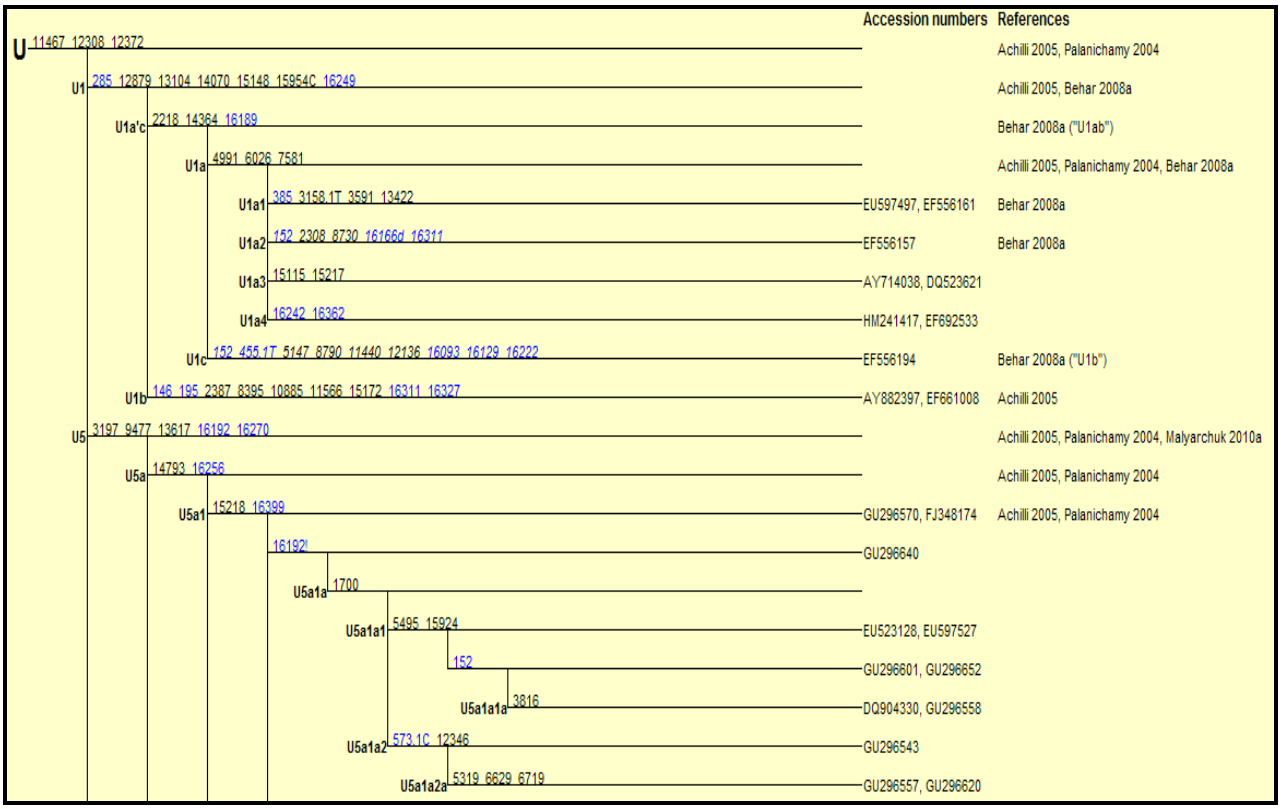


Fig. 4.26 Polimorfismi specifici per il sottoaplogruppo U5a1a1

Altri 2 campioni erano ascrivibili al sottoaplogruppo U3a, mentre altri 4 campioni erano ascrivibili ad altrettanti sottoaplogruppi (vedere tabella). L'estensione dell'analisi all'intero genoma mitocondriale ha consentito una migliore definizione dei sottoaplogruppi U5a1 in 2 U5a1a1, come detto sopra, 1 U5a1a2a ed infine un quarto come U5a1f (Palanichamy 2004; Achilli et al, 2005; Malyarchuk et al, 2010a). Relativamente al sottoaplogruppo U3a, uno dei due campioni è stato ulteriormente definito come U3a1 (Achilli et al, 2005; Metspalu et al, 2006; Behar et al, 2008a). Degli altri 4 campioni, ascrivibili ad altrettanti sottoaplogruppi, soltanto uno, U4a2, è rimasto definito già dall'analisi del D-loop, gli altri 3 sono stati rifiniti dai polimorfismi del codificante (vedere tab. IV.20) (Achilli et al, 2005; Metspalu 2006; González et al, 2006; Malyarchuk et al, 2008a). Vale la pena evidenziare, tra questi ultimi, la presenza

dell'aplogruppo U6, relativo al campione GEA87 (Isole), per il quale l'ulteriore definizione ha portato a classificarlo come **U6a7a1**, essendo presenti tutti i polimorfismi che lo definiscono, eccetto il polimorfismo C16278T. Anche in questo caso è plausibile una retromutazione verso la forma C16278C, trattandosi, questo, di un *hot spot* mutazionale (vedere fig. 4.27) (Olivieri et al, 2006).

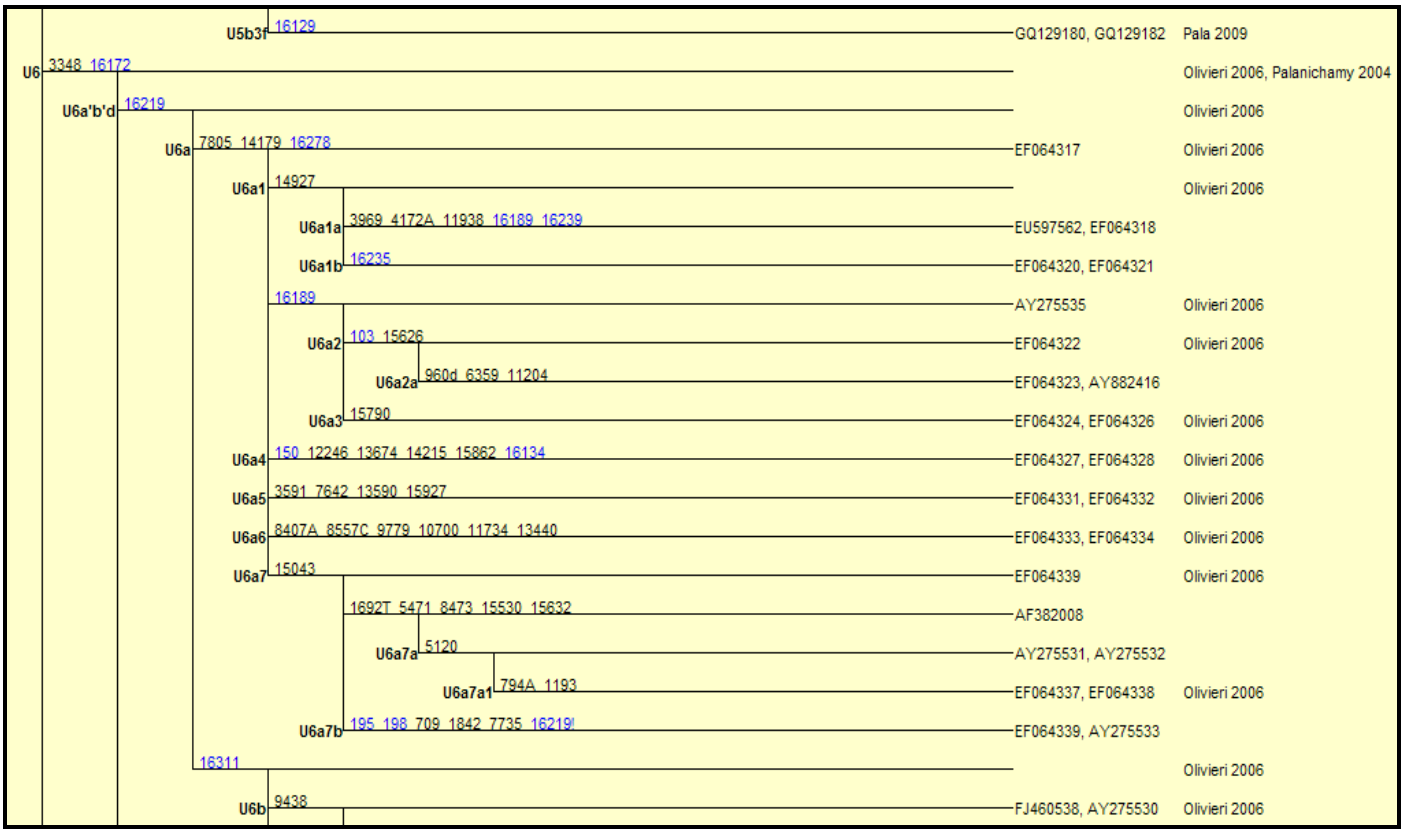


Fig. 4.27 Polimorfismi specifici per il sottoaplogruppo U6a7a1

Non è fuori luogo ricordare, infine, che l'aplogruppo U6 è di origine africana, proveniente, a sua volta ed al pari del sottoaplogruppo africano M1 (González et al, 2007), dai "Corridori Levantini" del Sudest asiatico (Olivieri et al, 2006) (fig. 4.28).

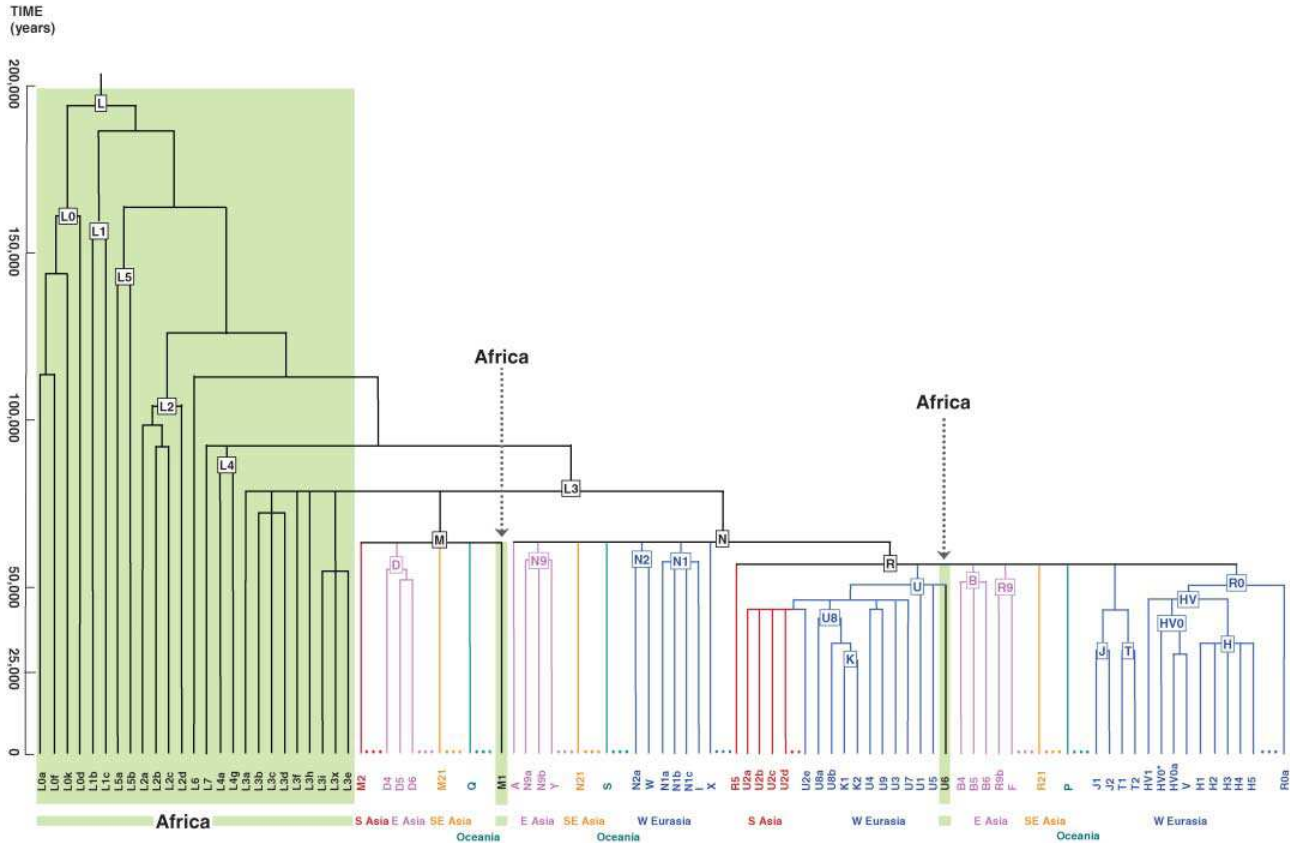


Fig. 4.28 Sono indicati in verde gli aplogruppi africani, ed in altri colori quelli relativi ad altre zone (Olivieri et al, 2006).

Gli aplogruppi Caucasici J e T sono aplogruppi fratelli, ognuno dei quali presente al 10% circa all'interno delle popolazioni caucasiche di origine europea e statunitense (Torroni et al, 1996; Macaulay et al, 1999a; Allard et al, 2002). Il cluster J/T è raggruppato dai polimorfismi diagnostici della regione codificante T4216C, A11251G e C15452A e dal polimorfismo diagnostico di HVI T16126C (Torroni et al, 1994; Torroni et al, 1996; Macaulay et al, 1999a; Finnila et al, 2001; Palanichamy et al, 2004; Derenko et al, 2007; Logan, 2009). L'aplogruppo J è caratterizzato dalle mutazioni nella regione codificante A10398G, A12612G e G13708A e dalle mutazioni diagnostiche HVI/HVII C16069T e C295T (Finnila et al, 2001; Herrnstadt et al, 2002; Allard et al, 2002;

Palanichamy et al, 2004; Derenko et al, 2007; Logan, 2009). L'aplogruppo T è caratterizzato dalle mutazioni specifiche della regione codificante G709A, G1888A, A4917G, G8697A, T10463C, G13368A, G14905A, A15607G e G15928A, e dal polimorfismo specifico di HVI C16294T (Finnila et al, 2001; Herrnstadt et al, 2002; Palanichamy et al, 2004; Kivisild et al, 2006). Nel presente studio sono stati individuati 2 campioni appartenenti all'aplogruppo J e 4 al T. I polimorfismi del D-loop hanno permesso di classificare i 2 campioni J come appartenenti ai sottoaplogruppi J1c2 e J1d; quest'ultimo è stato ulteriormente definito come J1d1 dal completamento della sequenza. Per quanto riguarda l'aplogruppo T, i relativi campioni sono stati tutti classificati completamente dai marcatori della regione di controllo (Richards et al, 2000; Palanichamy et al, 2004; Kivisild et al, 2006). Si nota, in particolare, la presenza di due soggetti (GEA107 e GEA108), entrambi del Nord Italia, identici al 100% nel loro genoma mitocondriale: ciò lascia supporre che i medesimi abbiano un comune ascendente in linea materna.

Il macroaplogruppo N è caratterizzato dai polimorfismi A8701, T9540 e T10873, che sono mutazioni antiche e stabili; W ed X sono due suoi sottoaplogruppi.

Al sottoaplogruppo N1b è ascrivibile il campione GEA70, ipotizzando una retromutazione nella posizione 16519 (Palanichamy et al, 2004; Behar et al, 2006; Derenko et al, 2007). C'è da tener presente, comunque, che tale posizione è un forte *hot spot* mutazionale e nemmeno viene considerata nella costruzione di parecchi alberi filogenetici, incluso quello cui ci si sta riferendo (van Oven M e Kayser M. 2008). Il

completamento della sequenza ha definito il campione come N1b1d (Feder et al, 2008). E' stato poi individuato un campione appartenente all'aplogruppo W, ma classificabile come "intermedio" (GEA85, vedere sopra) e non risolto dai polimorfismi del DNA codificante (Finnilä et al, 2001; Palanichamy et al, 2004; Kivisild et al, 2006; Metspalu et al, 2006; Behar et al, 2008a). Il sottoaplogruppo **X2** è rappresentato dal campione GEA59 (fig. 4.29), mancante però del polimorfismo A153G. Anche in questo caso è plausibile una retromutazione verso la forma A153A, trattandosi, questo, di un *hot spot* mutazionale (Reidla et al, 2003; Kivisild et al, 2006; Derenko et al, 2007; Tamm et al, 2007; Achilli et al, 2008; Perego et al, 2009).

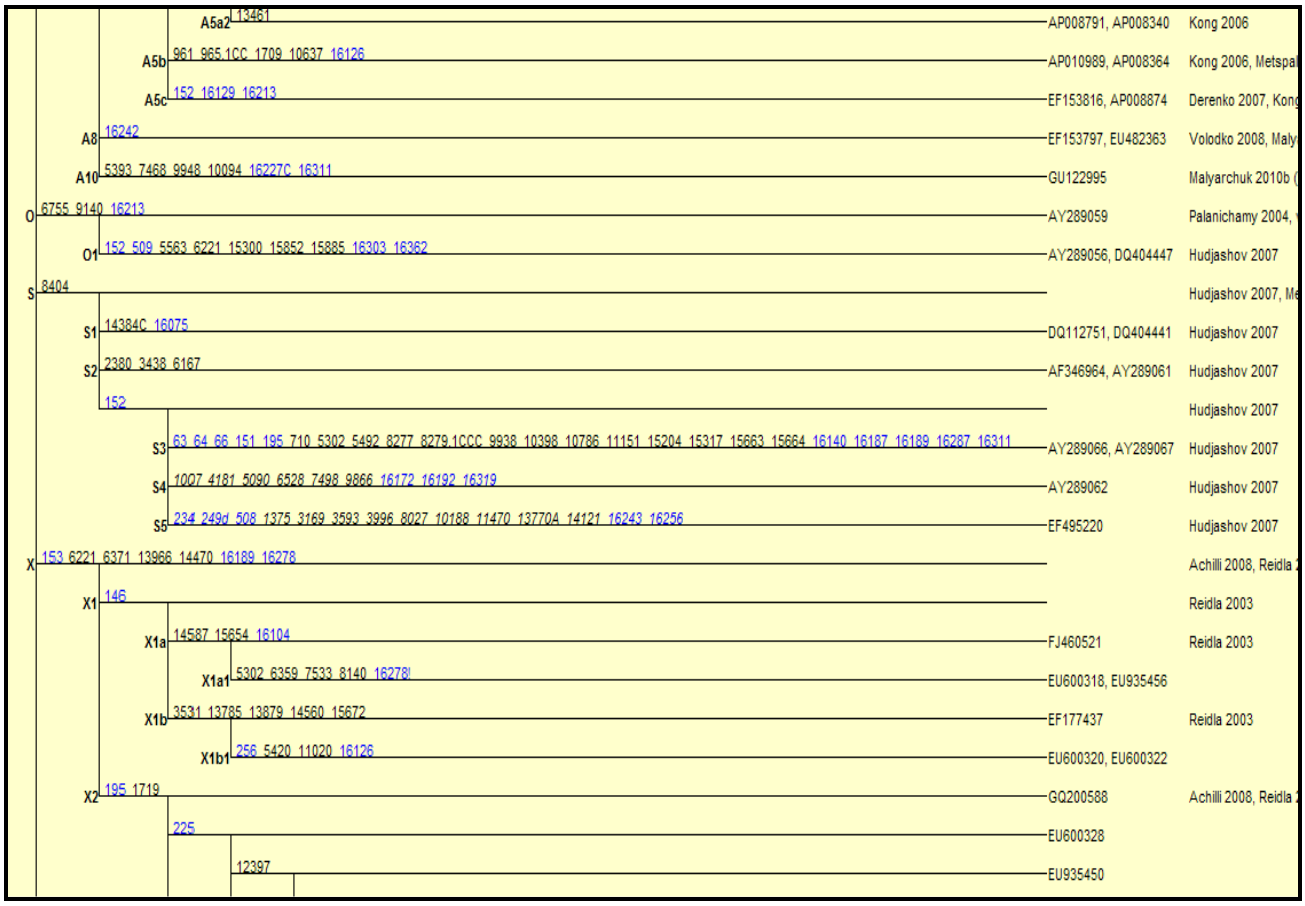


Fig. 4.29 Polimorfismi specifici per il sottoaplogruppo X2

L’aplogruppo L3 è tipicamente di orgine africana e rappesenta l’origine di tutti gli altri (vedere Introduzione). A questo specifico aplogruppo è stato ascritto il campione GEA63 (Sud Italia), come L3d relativamente al D-loop, ed L3d1b1 a sequenza completata (Behar et al, 2008b; Malyarchuk et al, 2008b).

Le tabelle IV.22 e IV.23, in cui si tiene conto solo degli aplogruppi principali, riassumono rispettivamente le osservazioni relative alla frequenza degli aplogruppi sul totale dei campioni e in base alla suddivisione in zone.

Aplogruppi principali	Frequenza nei campioni GEA	Frequenza nei caucasici
H	45%	40.9%
W	2.5%	2.1%
X	2.5%	4.1%
U	25.0%	14.3%
J	5%	10.5%
T	10%	12.7%
K	0%	8.0%
V	5%	3.2%
Altri	5%	2.1%

Tab. IV.22 Confronto delle frequenze degli aplogruppi principali osservati nel nostro campione della popolazione italiana e le frequenze dei caucasici (Torroni et al, 1996)

Sebbene la dimensione del campione studiato sia piccola, le percentuali osservate sono simili a quelle delle popolazioni caucasiche: l’aplogruppo H, con i suoi sottoaplogruppi, è il più rappresentato (quasi il 50%), seguito dagli aplogruppi U, T e J. Il nostro studio conferma precedenti osservazioni ricavate da studi di popolazione in Italia (Torroni et al, 1996; Bini et al, 2008; Brisighelli et al, 2008; Grignani et al, 2009). Infine, anche per quanto riguarda l’analisi territoriale, il numero dei campioni analizzati, nonché

la loro ineguale ripartizione geografica, non consente un'analisi statistica accurata. Tuttavia, i dati sembrano indicare un aumento della frequenza degli aplogruppi rari nel Sud Italia (vedere anche tabella riassuntiva IV.24).

Zona	Frequenza aplogruppo H nel campione GEA	Frequenza aplogruppo H nella popolazione italiana
NORD	**55.5%**	**46.9%**
CENTRO	**55.5%**	**35.6%**
SUD	**37.5%**	**37.9%**
ISOLE	**33.3%**	**45.7%**

Tab. IV.23 Confronto della frequenza dell'aplogruppo H osservata nelle diverse aree geografiche tra i 40 mtDNA GEA e le frequenze riportate in letteratura (Achilli et al, 2004).

Zona	Campione	Aplogruppo di appartenenza
NORD	GEA65 (12)	**H**
	GEA73 (13)	**H1c**
	GEA76 (12)	**H**
	GEA79 (30)	**U5a1f**
	GEA82 (14)	**H5a1**
	GEA86 (9)	**H**
	GEA107 (38)	**T2b3**
	GEA108 (38)	**T2b3**
	GEA111 (33)	**U3a**
CENTRO	GEA28 (28)	**U5a1a1**
	GEA44 (16)	**V**
	GEA80 (35)	**U4c1a**
	GEA81 (14)	**H35**
	GEA88 (15)	**H10a1**
	GEA98 (9)	**H2b**
	GEA100 (9)	**H1**
	GEA104 (15)	***H11'12***
	GEA105 (32)	**U3a1**
SUD	GEA57 (11)	**H15b**
	GEA58 (8)	**H33**
	GEA59 (34)	**X2**
	GEA60 (15)	**H13a1a1a**
	GEA61 (37)	**T1a**
	GEA62 (30)	**U5a1a2a**
	GEA63(32)	**L3d1b1**
	GEA64 (38)	**T2**
	GEA66 (8)	**H33**
	GEA67 (33)	**J1c2**
	GEA68 (29)	**U5a1a1**
	GEA69 (14)	***H1a-b'f'g'k'q***
	GEA70 (41)	**N1b1d**
	GEA71 (14)	**H**
	GEA72 (36)	**U8b1**
	GEA109 (36)	**J1d1**
ISOLE	GEA21 (17)	**H13a2a**
	GEA83 (7)	**H2a2b1**
	GEA84 (29)	**U4a2**
	GEA85 (32)	***W3'6***
	GEA87 (29)	**U6a7a1**
	GEA96 (17)	**V**

Tab. IV.24 Schema riassuntivo degli aplogruppi individuati in ogni campione: nella prima colonna è riportata la ripartizione in zone della penisola italiana, mentre a fianco ad ogni campione viene indicato il numero di polimorfismi riscontrato durante l'analisi. Per gli aplogruppi evidenziati in fucsia, vedere legenda tab. IV.21.

4.6.1 Individuazione di SNPs utili: quali risvolti?

Data la natura della variazione all'interno del mtDNA, basata principalmente su polimorfismi da mutazioni puntiformi, il sistema mitocondriale si presta particolarmente ai saggi di SNPs (Vallone et al, 2004). Ciò vale particolarmente nella regione codificante, nella quale i polimorfismi sono ampiamente dispersi. Tale dispersione rende il sequenziamento totale poco pratico, specialmente nel caso di campioni degradati, e poco gradevole nel momento in cui rivelasse mutazioni significativamente rilevanti da un punto di vista medico. Scopo principale del lavoro è stato quello di individuare potenziali siti di SNPs da proporre per saggi forensi immediati, su campioni altamente problematici, al fine di risolvere individui Caucasici che condividono il medesimo HVI/HVII. Criteri di massima sono stati: a) neutralità degli SNPs; b) condivisione; c) non ridondanza. Relativamente al primo, nel caso si volessero scartare tutti quelli che potenzialmente potrebbero indurre delle variazioni fenotipiche, si dovrebbe comunque fare attenzione a quelli che invece si evidenziano in un solo individuo: questi potrebbero essere, infatti, "mutazioni private", rare o uniche nella popolazione in esame. Queste, pertanto, costituirebbero una scelta non adeguata per degli SNPs da inserire in saggi di discriminazione forense. In merito alla condivisione, si è scelto di osservare, nell'insieme, tutti i polimorfismi ricavati dall'analisi ed inseriti nella definizione degli aplogruppi, al fine di individuare quelli realmente discriminanti. La non ridondanza è sicuramente il criterio più difficoltoso, trattandosi di una molecola chiusa e legata. La soluzione consiste nell'individuare siti particolarmente variabili nell'ambito di tipi comuni HVI/HVII. A mero titolo di esempio, si può ricordare la già citata posizione

16519, uno dei siti più veloci evolutivamente (Chen et al, 1995; Richards et al, 1998; Finnila et al, 2001; Ingman e Gyllesten, 2001).

Lo studio ha evidenziato diverse categorie di polimorfismi: marcatori sull'albero filogenetico (gli SNPs già propriamente definiti: vedere anche tabella IV.25), polimorfismi più o meno nuovi e mutazioni private. Nell'ambito di queste tre categorie è possibile definire ulteriori sottocategorie, come pure sottocategorie che derivano dall'intersezione delle prime tre.

a) Per quanto riguarda gli **SNPs**, ve ne sono alcuni di antica definizione ed altri che sono stati definiti più recentemente, man mano che ha preso piede il sequenziamento di tutto il genoma mitocondriale, l'implementazione della banca dati (Behar et al, 2008) e l'affinamento nella definizione degli aplogruppi (vedere prospetto e bibliografia relativa ai vari aplogruppi). SNPs "vecchi" e "nuovi", da soli o in combinazione tra di loro (vedere prospetto dettagliato, Tabb. IV.7 → IV.13), sono stati anche osservati: 1) in associazione a varianti non pubblicate, non depositate in Gene Bank; 2) in associazione a sospetta variazione fenotipica; 3) in associazione a fenotipi patologici; 4) come mutazioni private in altri aplogruppi e/o sottoaplogruppi, anche associate ai casi precedenti, così come pure alcune varianti non pubblicate si sono ritrovate associate a SNPs con reports di associazione a patologie. Per il dettaglio relativo alle varianti non pubblicate, le variazioni fenotipiche e patologiche si rimanda al capitolo relativo all'analisi dei polimorfismi. Qui preme sottolineare come l'eterogeneità dei polimorfismi individuati rifletta l'enorme variabilità del DNA mitocondriale,

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16298C**	**T72C**									
GEA44												
V	***Coding Region***	A1438G	A2706G	**G4580A**	A4769G	C7028T	**G11719G**	**C14766C**	**C15904T**			
	Control Region	**T16298C**	**T72C**									
GEA96												
V	***Coding Region***	A1438G	A2706G	**G4580A**	A4769G	C7028T	**G11719G**	**C14766C**	**C15904T**			
	Control Region	T16519C	**A73A**									
GEA65												
H	***Coding Region***	A1438G	**A2706A**	A4769G	**C7028C**							
	Control Region	T16519C	**A73A**									
GEA71												
H	***Coding Region***	A1438G	**A2706A**	A4769G	**C7028C**							
	Control Region	T16519C	**A73A**									
GEA76												
H	***Coding Region***	A1438G	**A2706A**	A4769G	**C7028C**							
	Control Region	T16519C	**A73A**									
GEA86												
H	***Coding Region***	A1438G	**A2706A**	A4769G	**C7028C**							

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	T16519C	**A73A**									
GEA100												
H1	***Coding Region***	A1438G	**G3010A**	A4769G								
	Control Region	**T16189C**	T16519C	**A73A**								
GEA69												
H1a-b'f'g'k'q	***Coding Region***	A1438G	**G3010A**	A4769G								
	Control Region	T16519C	**T477C**									
GEA73												
H1c	***Coding Region***	A1438G	**G3010A**	A4769G								
	Control Region	**A16235G**	**C16291T**									
GEA83												
H2a2b1	***Coding Region***	**A1438A**	**A4769A**									
	Control Region	**T16311C**	**T152C**									
GEA98												
H2b	***Coding Region***	**A1438A**	A4769G	**T8598C**								
	Control Region	**T16304C**	**C456T**									
GEA82												
H5a1	***Coding Region***	A1438G	**T4336C**	**C15833T**								

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16114T**	T16519C									
GEA88												
H10a1	***Coding Region***	A1438G	**T4216C**	A4769G	**T14470A**	**A14548G**						
	Control Region	T16519C	**A73A**	**T195C**								
GEA104												
H11'12	***Coding Region***	A1438G	A4769G									
	Control Region	**A73A**										
GEA60												
H13a1a1a	***Coding Region***	A1438G	**C2259T**	**A4745G**	A4769G	**G7337A**	**T13326C**	**C13680T**	**C14872T**			
	Control Region	T16519C	**A73A**									
GEA21												
H13a2a	***Coding Region***	**G709A**	**A1008G**	**A1438A**	**C2259T**	A4769G	**C14872T**					
Allele raro												
	Control Region	**T55C**	**T57G**									
GEA57												
H15b	***Coding Region***	A1438G	**T3847C**	A4769G	**T6253C**							
	Control Region	**A73A**										
GEA58												
H33	***Coding Region***	A1438G	A4769G	**G11447A**								

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**A73A**										
GEA66												
H33	***Coding Region***	A1438G	A4769G	**G11447A**								
	Control Region	T16519C	**A73A**									
GEA81												
H35	***Coding Region***	A1438G	**C3342T**	A4769G								
	Control Region	**C16192C**	**C16256T**	**C16270T**	**A16399G**	**A73G**						
GEA68												
	Coding Region	A1438G	**T1700C**	A2706G	**T3197C**	A4769G	**T5495C**	C7028T	**G9477A**	**A11467G**	**G11719A**	**A12308G**
U5a1a1		**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**	**A15924G**					
	Control Region	**C16192C**	**C16256T**	**C16270C**	**A16399G**	**A73G**						
GEA28												
U5a1a1	***Coding Region***	A1438G	**T1700C**	A2706G	**T3197C**	A4769G	**T5495C**	C7028T	**G9477A**	**A11467G**	**G11719A**	**A12308G**
Retromutazione		**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**	**A15924G**					
	Control Region	**C16192C**	**C16256T**	**C16270T**	**A16399G**	**A73G**	**573.1C**					
GEA62												
U5a1a2a	***Coding Region***	A1438G	**T1700C**	A2706G	**T3197C**	A4769G	**A5319G**	**A6629G**	**T6719C**	C7028T	**G9477A**	**A11467G**
		G11719A	**A12308G**	**C12346T**	**G12372A**	**T13617C**	**C14766T**	**A14793G**	**A15218G**			

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16192T**	**C16256T**	**C16270T**	**A16399G**	**A73G**						
GEA79												
U5a1f	***Coding Region***	A1438G	A2706G	**T3197C**	A4769G	**G6023A**	C7028T	**G9477A**	**A11467G**	**G11719A**	**A12308G**	**G12372A**
		T13617C	**C14766T**	**A14793G**	**A15218G**							
	Control Region	**T16172C**	**A16219G**	**C16278C**	**A73G**							
GEA87												
U6a7a1	***Coding Region***	**T794A**	**T1193C**	A1438G	**A1692T**	A2706G	**A3348G**	A4769G	**A5120G**	**G5471A**	C7028T	**G7805A**
Hot spot		**T8473C**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**A14179G**	**C14766T**	**G15043A**	**T15530C**	**C15632T**	
	Control Region	**A16343G**	**G16390A**	T16519C	**A73G**	**C150T**						
GEA111												
U3a	***Coding Region***	A1438G	**A1811G**	**A2294G**	A2706G	**T4703C**	A4769G	**C6518T**	C7028T	**G9266A**	**A10506G**	**A11467G**
		G11719A	**A12308G**	**G12372A**	**C13934T**	**A14139G**	**C14766T**	**T15454C**				
	Control Region	**A16343G**	**G16390A**	T16519C	**A73G**	**C150T**						
GEA105												
U3a1	***Coding Region***	A1438G	**A1811G**	**A2294G**	A2706G	**G3010A**	**T4703C**	A4769G	**C6518T**	C7028T	**G9266A**	**A10506G**
		A11467G	**G11719A**	**A12308G**	**G12372A**	**C13934T**	**A14139G**	**C14766T**	**T15454C**			
	Control Region	**T16356C**	T16519C	**A73G**	**T195C**	**T310C**	**G499A**					
GEA84												
U4a2	***Coding Region***	A1438G	**A1811G**	A2706G	**T4646C**	A4769G	**T5999C**	**A6047G**	C7028T	**C8818T**	**C11332T**	**A11467G**
		G11719A	**A12308G**	**G12372A**	**C14620T**	**C14766T**	**T15693C**					

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**C16179T**	**T16356C**	T16519C	**A73G**	**T195C**	**G499A**					
GEA80												
U4c1a	***Coding Region***	A1438G	**A1811G**	A2706G	**T4646C**	A4769G	**A4811G**	**T5999C**	**A6047G**	**A6146G**	C7028T	**T9070G**
		T10907C	**T11009C**	**C11332T**	**A11467G**	**G11719A**	**A12308G**	**G12372A**	**C14620T**	**C14766T**	**C14866T**	**T15693C**
	Control Region	**T16189C**	**C16234T**	T16519C	**A73G**	**T195C**						
GEA72												
U8b1	***Coding Region***	A1438G	**A1811G**	A2706G	**A3480G**	A4769G	C7028T	**G9055A**	**T9698C**	**A11467G**	**G11719A**	**A12308G**
		G12372A	**C14167T**	**C14766T**								
	Control Region	**C16069T**	**T16126C**	**A73G**	**G185A**	**A188G**	**G228A**	**C295T**	**C462T**			
GEA67		**T489C**										
J1c2												
	Coding Region	A1438G	A2706G	**G3010A**	**T4216C**	A4769G	C7028T	**A10398G**	**A11251G**	**G11719A**	**A12612G**	**G13708A**
		C14766T	**T14798C**	**C15452A**								
	Control Region	**C16069T**	**T16126C**	**C16193T**	**A16300G**	**A73G**	**T152C**	**C295T**	**C462T**	**T489C**		
GEA109												
J1d1												
	Coding Region	A1438G	A2706G	**G3010A**	**T4216C**	A4769G	C7028T	**G7789A**	**A7963G**	**A10398G**	**A11251G**	**G11719A**
		A12612G	**G13708A**	**C14766T**	**C15452A**							

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16126C**	**A16163G**	**C16186T**	**T16189C**	**C16294T**	**A73G**	T16519C				
GEA61												
T1a												
	Coding Region	**G709A**	A1438G	**G1888A**	A2706G	**T4216C**	A4769G	**A4917G**	C7028T	**G8697A**	**T10463C**	**A11251G**
		G11719A	**C12633A**	**G13368A**	**C14766T**	**G14905A**	**C15452A**	**A15607G**	**G15928A**			
	Control Region	**T16126C**	**C16294T**	**C16296T**	T16519C	**A73G**						
GEA64												
T2												
	Coding Region	**G709A**	A1438G	**G1888A**	A2706G	**T4216C**	A4769G	**A4917G**	C7028T	**G8697A**	**T10463C**	**A11251G**
		G11719A	**A11812G**	**G13368A**	**A14233G**	**C14766T**	**G14905A**	**C15452A**	**A15607G**	**G15928A**		
	Control Region	**T16126C**	**C16294T**	**C16296T**	**T16304C**	T16519C	**A73G**					
GEA108												
T2b3	***Coding Region***	**G709A**	**G930A**	A1438G	**G1888A**	A2706G	**T4216C**	A4769G	**A4917G**	**G5147A**	C7028T	**G8697A**
		T10463C	**A10750G**	**A11251G**	**G11719A**	**A11812G**	**G13368A**	**A14233G**	**C14766T**	**G14905A**	**C15452A**	**A15607G**
		G15928A										
	Control Region	**T16126C**	**C16294T**	**C16296T**	**T16304C**	T16519C	**A73G**					
GEA107												
T2b3	***Coding Region***	**G709A**	**G930A**	A1438G	**G1888A**	A2706G	**T4216C**	A4769G	**A4917G**	**G5147A**	C7028T	**G8697A**
		T10463C	**A10750G**	**A11251G**	**G11719A**	**A11812G**	**G13368A**	**A14233G**	**C14766T**	**G14905A**	**C15452A**	**A15607G**
		G15928A										

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**G16145A**	**C16176G**	**C16223T**	**G16390A**	**A73G**	**T152C**	**G185A**	**A188G**			
GEA70												
N1b1d												
	Coding Region	A1438G	**G1598A**	**C1703T**	**G1719A**	**C2639T**	A2706G	**C3921A**	A4769G	**C4960T**	**G5471A**	C7028T
		G8251A	**C8472T**	**A8701A**	**T8763C**	**A8836G**	**C9335T**	**T9540T**	**T10238C**	**A10398A**	**T10873T**	**A11362G**
		G11719A	**G12501A**	**C12705T**	**A12822G**	**C14766T**	**G15301G**					
	Control Region	**C16223T**	**C16292T**	T16519C	**A73G**	**A189G**	**C194T**	**T195C**	**T204C**	**G207A**		
GEA85												
W3'6												
	Coding Region	**G709A**	**T1243C**	A1438G	A2706G	**A3505G**	A4769G	**G5046A**	**G5460A**	C7028T	**G8251A**	**G8994A**
		C11674T	**G11719A**	**A11947G**	**T12414C**	**C12705T**	**C14766T**	**G15884C**				
	Control Region	**T16189C**	**C16278T**	T16519C	**A73G**	**A153A**	**T195C**					
GEA59												
X2		A1438G	**G1719A**	A2706G	A4769G	**T6221C**	**C6371T**	C7028T	**G11719A**	**C12705T**	**A13966G**	**T14470C**
Hot spot	***Coding Region***	**C14766T**										

Campione	Regione esaminata	Polimorfismi individuati										
	Control Region	**T16124C**	**C16223T**	T16519C	**A73G**	**T152C**	**(523d-524d)**					
GEA63												
L3d1b1	***Coding Region***	**T921C**	A1438G	A2706G	A4769G	**G5046A**	**G5147A**	**A6272G**	**T6680C**	C7028T	**A7424G**	**T8618C**
		A8701G	**T9540C**	**A10398G**	**T10873C**	**G11719A**	**C12705T**	**A13105G**	**T13886C**	**C14284T**	**C14766T**	**G15301A**

*eteroplasmia

Tab. IV.25 Individuazione degli aplogruppi nei 40 campioni esaminati. In verde, i polimorfismi del D-loop, già individuati come inclusi nei *mutation motifs*; in azzurro gli ulteriori polimorfismi, inclusi nell'albero filogenetico ed utilizzati per l'assegnazione definitiva ai vari aplogruppi (indicati sotto ogni campione); in fucsia le possibili retromutazioni (vedere testo); in grigio i polimorfismi la cui assenza determina l'assegnazione ad un dato aplogruppo: a titolo esemplificativo ne vengono indicati solo alcuni, sottintendendone la presenza nei relativi sottoaplogruppi e/o aplogruppi correlati: per ulteriori dettagli si rinvia al testo; in carattere normale i polimorfismi non discriminanti o già abbondantemente descritti (vedere anche tab…).

nonostante le ridotte dimensioni del campione esaminato; le implicazioni di ciascun polimorfismo, come sottolineato più volte, sono eterogenee a seconda del contesto. Tra gli esempi in merito possiamo citare il campione GEA59, ascrivibile all'aplogruppo X2, che contiene nella sua sequenza dei polimorfismi tipici di alcuni suoi sottoaplogruppi: G13708A, del complesso X2b'd e A6791G e T8503C, tipici di X2d: il numero di passaggi che sarebbe richiesto per arrivare a definire il campione come appartenente a tale sottoaplogruppo sarebbe eccessivo e richiederebbe molti eventi evolutivi, per cui è più probabile che si tratti di mutazioni private (metodo della massima parsimonia). Un altro esempio per tutti riguarda il già ricordato T16519C: pur presente in vari *mutation motifs*, cui si rimanda (Tab. IV.18), lo si è riscontrato in alcuni aplogruppi nei cui *mutation motifs* non era incluso (H,H1, H1a-b'f'g'k'q, H10a1, H11'12 H13a2a, H35, U3a, U3a1, U8b1, L3d1b1), e, al contrario, non lo si è ritrovato in N1b, nella cui stringa era incluso. Questa categoria di SNPs è quella che, più delle altre due, soddisfa il criterio della condivisione, essendo stati già individuati come specifici marcatori sull'albero filogenetico. Il criterio, invece, della neutralità, lascia aperte molte questioni. E' difficile, se non impossibile, attribuire un effetto patologico ad un polimorfismo riportato negli studi popolazionistici (Accetturo et al, 2006), specialmente se questo è molto frequente nella popolazione o ricorrente nell'evoluzione: pertanto il dato relativo alla frequenza di popolazione è un'informazione essenziale per dare come certo un effetto negativo di quel dato polimorfismo. Nella casistica indicata sono presenti anche marcatori "storici", per i quali è stato formulato il sospetto di variazione fenotipica, o casi di marcatori associati a patologie ed ormai consolidati, o ancora casi di marcatori che cadono in regioni

particolarmente delicate, quali quelle degli RNA: basta, per tutti, citare il polimorfismo nella posizione 3010 (eccezione al criterio “conservativo” per la selezione dello SNP, trattandosi comunque di un nucleotide del 16S rRna che, al pari di quelli appartenenti ai tRna, sarebbe consigliabile evitare di eleggere come SNP): si tratta, infatti, di una mutazione ben caratterizzata, di ampio riscontro in letteratura, e non sospettata, allo stato attuale, di essere associata ad una qualche condizione patologica (Mehta et al, 1989; Finnila et al, 2001). Infine si è rivelata molto utile nella discriminazione forense all’interno dell’aplogruppo H. Questa ampia variazione della 3010 nella popolazione fa pensare che il polimorfismo sia governato dalla teoria Neutrale dell’Evoluzione Molecolare (Kimura, 1983). In generale, però, tali attribuzioni hanno importanti risvolti, oltre che da un punto di vista forense, anche da un punto di vista clinico e, come già ricordato nella parte introduttiva, etico: molte associazioni delle varianti del mtDNA sono state infatti testate in molti caratteri complessi, quali la longevità (De Benedictis et al, 1999; Bilal et al, 2008), il morbo di Alzheimer (Elson et al, 2006) e l’infertilità maschile (Pereira et al, 2007), solo per citarne alcuni. Le informazioni presentate in questo lavoro sono state meticolosamente raccolte facendo riferimento a MITOMAP, lo storico portale sulla variabilità del DNA mitocondriale (Ruiz-Pesini et al, 2007), utile soprattutto per definire la novità di un dato polimorfismo, nonché al mtDB – Human Mitochondrial DNA Database (Ingman e Gyllensten, 2006), utile riferimento per la frequenza, la localizzazione ed il tipo di mutazioni osservate (se sinonime o non sinonime, nel caso di geni codificanti proteine). L’unione delle informazioni relative alle mutazioni patogeniche ed ai polimorfismi degli aplogruppi non ha dato dei risultati

chiari, come atteso per un campione ristretto e per uno studio che, per la sua complessità, facilmente può andare incontro ad errori (Bandelt et al, 2008): un aiuto per poter formulare qualche ipotesi in più è arrivato dal continuo confronto con i dati presenti in Gene Bank, che a buon motivo può essere vista come un database dinamico e sicuramente aggiornato. I due database sopra citati, per quanto aggiornati, poco comunicano circa la dinamica dei polimorfismi e la variazione delle popolazioni, desumibile solo dai dati completi che vengono via via inseriti. L'importanza di fare riferimento alle frequenze del polimorfismo in un database dinamico, al fine di estrapolarne un eventuale effetto patologico od una neutralità che ne consenta un'inclusione sicura nei saggi forensi, viene evidenziata dai dati riportati nel prospetto, relativi alla seconda categoria di polimorfismi.

b) I **polimorfismi nuovi**. Tra questi ve ne sono alcuni descritti in mtDB e con riscontri in Gene Bank, che già sono stati classificati come SNPs discriminanti ed inseriti nell'albero filogenetico; per altri, invece, parimenti descritti, tale inserimento non è ancora avvenuto. Per questa, come per la precedente categoria di SNPs, oltre alle frequenze nelle popolazioni globali e locali è bene considerare: la conservazione del polimorfismo anche tra specie diverse (vedere la discussione in merito ai tRna), il già citato effetto della mutazione nelle proteine, cioè se essa è sinonima o meno, ed il contesto degli altri siti polimorfici nella quale la medesima si trova. Studi recenti (Pereira et al, 2009) hanno ad esempio dimostrato che un ristretto gruppo di variazioni non sinonime è meglio tollerato e quindi tali variazioni sono relativamente ben rappresentate nella popolazione esaminata (oltre 5000 soggetti): tra queste, spiccano

quelle che interessano la treonina, dato peraltro evidenziabile anche in questo studio, nonostante l'esiguità del campione. Per alcune mutazioni, infatti, si è ipotizzato che fossero compensative. Per gli RNA non si è trascurato di considerare le strutture secondarie, specie relativamente ai nuovi polimorfismi, le cui caratteristiche sono critiche per poter dedurre eventuali effetti patologici di determinate variazioni. Nell'ambito di questa categoria di polimorfismi ve ne sono infine alcuni, assenti in entrambi i data base di riferimento ma con qualche riscontro in Gene Bank ed altri di assoluto, nuovo riscontro, nel presente studio (vedere tabella riassuntiva 4.26). Su questi come su quelli appena ricordati, non ancora validati come SNP, sarà necessario proseguire gli studi nel senso di cercarne un più ampio riscontro nel database di popolazione, una volta che questo venga ampliato, e di valutarne gli eventuali aspetti patologici alla luce dei criteri sopra esposti.

Variazione	S/NS	Gene	Osservazioni	Campioni
A1636G	--	$tRna^{Val}$	**Ansa contenente l'anticodone**	109
T6512C	S	COI		87
T8945G	NS	ATP6	? (stop-codon)	21
A4021T	NS	ND1		72
C5345T	S	ND2		61
C5424T	NS	ND2		111
C11356T	S	ND4		61
A13866G	S	ND5		70

Tab. IV.26 I polimorfismi di nuova descrizione evidenziati nel presente studio.

c) **Mutazioni private**. Per quanto singole, associate a sospetta variazione fenotipica o a chiara condizione patologica, anche per queste vale quanto detto sopra. Tra le mutazioni private riscontrate nello studio, ve ne sono alcune condivise: la A9115G, condivisa da due soggetti di aplogruppo H33: per questi, come già detto, non si può escludere un rapporto di parentela in senso stretto (es. fratelli), ma non avendo informazioni in tal senso si può ipotizzare, più genericamente, un ascendente comune in linea materna. Stesso discorso vale per i seguenti polimorfismi: C4843T, A9007G e G10993A, condivisi da due soggetti di sottoaplogruppo T2b3. Per quanto possa essere ipotizzabile una loro inclusione nell'albero filogenetico, ad ulteriore definizione dei relativi aplogruppi, tutti e 4 hanno pochissimi hits (da 2 a 8) in Gene Bank e sono presenti in popolazioni eterogenee: molto più probabilmente trattasi di mutazioni private condivise a livello di ascendenza e presenti in varie etnie.

Per quanto concerne, infine, le *varianti non pubblicate*, variamente presenti nel campione in esame, maggiore informazione dovrebbe essere data circa l'assetto complessivo del genoma mitocondriale: la descrizione di un polimorfismo in un soggetto affetto da una data patologia, non necessariamente indica un rapporto causa-effetto tra i due! L'annotazione della sequenza complessiva e la sua deposizione in Gene Bank possono rendere più informativa l'osservazione (ovviamente previo consenso informato e con garanzia di anonimato, vedere Introduzione). Quello che traspare dal presente studio

è che sembra esserci una maggiore frequenza di determinate patologie, descritte contestualmente a determinati polimorfismi, in specifiche etnie (es. tumore della bocca negli Indiani Asiatici, adenoma sporadico delle paratiroidi negli USA, e così via), ma altre condizioni, ad esempio legate alle abitudini di vita, non possono essere escluse come cause concomitanti nel determinare tali fenotipi.

Come evidente dai dati finora discussi, vi sono notevoli differenze, tra i vari aplogruppi, nel rapporto tra mutazioni sinonime e non; è stato visto che tale variazione segue anche quella delle diverse aree geografiche, alle diverse latitudini (Wiesbauer et al, 2006). Differenti linee mitocondriali presentano differente funzionalità, e questo può incidere sulle manifestazioni patologiche (Moilanen et Majamaa, 2003; Ghezzi et al, 2005; Marcuello et al, 2008; Suissa et al, 2009). Ad esempio, associazioni positive e negative degli aplogruppi sono state riscontrate nella astenozoospermia, nella sclerosi multipla, ma anche nella sopravvivenza post-sepsi e nella longevità (Wei et al, 2008; Martínez-Redondo et al, 2010), significando che gli aplogruppi, inizialmente ritenuti fenotipicamente indistinguibili, possono essere determinanti nelle malattie e/o nei caratteri multifattoriali. Tuttavia molte lacune devono ancora essere colmate: la maggior parte di questi dati deriva infatti da sottopopolazioni di specifiche aree geografiche (vedasi le varianti individuali, di cui sopra) e comunque di popolosità bassa. Per essere significativi e definitivi, tali studi richiedono un adeguato numero di soggetti, da centinaia a migliaia, da inserire in confronti caso-controllo (Abu Amero et al, 2006).

4.6.1.1 Un caso problematico: il campione GEA21

Nel corso di questo studio ci si è imbattuti in difficoltà sia di natura tecnica che interpretativa. Si è già detto delle problematiche di sequenziamento relative al D-loop, risolte con primers interni, nonché di altre zone del DNA codificante, per le quali si è reso necessario dover allestire degli esperimenti di PCR a gradiente per settare nuove condizioni di amplificazione. Una problematica tuttora rimasta aperta riguarda il sequenziamento, per alcuni campioni, della regione coperta dalla coppia di primer 8FR; gli studi sono tuttora in corso, per comprendere le basi teoriche di tale problematicità. Sta di fatto che per GEA21 non è stato possibile ottenere tale sequenza con nessuna delle metodiche in questione. Si è riusciti ad ottenere l'amplificato 7F – 9R (vedere Materiali e Metodi), ma non l'amplificato 8FR ordinario e nessuno dei due relativi prodotti di sequenza, se non nel primissimo tratto, che si sovrappone alla parte finale dell'amplicone 7FR. In corrispondenza di tale tratto vi è l'ultima porzione del gene ND2, mentre la rimanente porzione dell'8FR interessa i tRna per il Triptofano, l'Alanina, l'Asparagina, la Cisteina e la Tirosina, la cui ultima porzione è risolta dalla sovrapposizione con l'inizio dell'amplicone 9FR. Contestualmente, nel medesimo campione, è stata descritta una mutazione non sense, nuova, che trasforma un aminoacido in un codone di stop (vedere analisi dei polimorfismi. Come detto, tali mutazioni hanno netta valenza patologica; inoltre, riferendosi alla letteratura corrente, si osserva che le mutazioni patologiche descritte a livello del gene ATP6 (8527 – 9207) spesso correlano con mutazioni a carico di diversi tRna (Carrozzo et al, 2006; Ahari et al, 2007; Kasraie et al, 2008; Safaei et al, 2009) che, come detto più volte, sono i "depositari" delle mutazioni più pericolose. L'impossibilità di sequenziare alcuni di questi, associata a questa

particolare mutazione di stop, lascia ipotizzare una carenza a livello della funzionalità di ATP6. A ciò si aggiunge l'allele raro A1438A, per il quale supponiamo una retromutazione nel contesto dell'aplogruppo di appartenenza di GEA21, descritto in associazione a disordini psichiatrici (Rollins et al, 2009). Nell'insieme delle mutazioni descritte per questo soggetto, ve ne è una, la 12662, individuata unicamente negli isolati genetici della Sardegna (Fraumene et al, 2006); l'appartenenza del soggetto all'Italia insulare potrebbe ricondurlo a tali popolazioni. Non si dispone di informazioni circa la provenienza esatta del soggetto, né circa le sue condizioni patologiche; quanto detto finora, però, pone un quesito importante (Logan, 2005), dagli ampi risvolti.

4.6.2 Quale database mitocondriale?

Abbiamo visto che la ricerca di SNPs mette lo scienziato forense di fronte a difficoltà tecniche, interpretative ed etiche; nell'ottica di costruire/implementare un data base mitocondriale, appare chiaro che diversi fattori devono essere tenuti in considerazione:

a) Tipologia di campioni: non vi è rigidità di criterio, possono essere inclusi soggetti volontari del laboratorio forense (normalmente già tipizzati per il DNA nucleare), eventuali donatori di sangue di concerto con una banca del sangue, vittime ed indagati provenienti da casi reali. Se possibile deve essere recuperata l'informazione circa l'ascendenza in via materna, almeno fino a due generazioni indietro; se non è possibile, specie per gli immigrati, è preferibile non includerli nel data base, o quanto meno, segnalare tale assenza di informazione. Circa

l’annotazione specifica, relativa all’etnia, si può fare riferimento alle linee guida formulate dall’EMPOP (Parson e Dür, 2007).

b) Sequenziamento nell’ambito del D-loop (auspicabile): tradizionalmente limitato all’HVI, poi esteso, più recentemente, anche all’HVII (nella letteratura si parla infatti di “HVI/HVII common types”, e da qualche laboratorio, esteso anche all’HVIII (Bini et al, 2003), è comunque da ritenersi superato (Carracedo et al, 2010). Per quanto tale regione contenga, al suo interno, dei siti inclusi nelle stringhe degli aplogruppi Eurasiatici (Achilli et al, 2005; Kong et al, 2003; Palanichamy et al, 2004; Sun et al, 2006), cui pure si è fatto riferimento nel presente studio, è comunque certo che range troppo stretti di analisi vanno ad escludere un certo numero di varianti nucleotidiche frequenti che sono tipiche di diversi aplogruppi, come la transizione in 16051 (ad es. nell’aplogruppo U2), 16390 (L2, N1b1, Uea ecc), 16399 (L2d, U5A1), 72 (HVO), e così via. Il nostro studio, in tal senso, già si pone in avanti, avendo voluto utilizzare, come primo screening, quello dato dal sequenziamento dell’intera regione di controllo. Gli stessi mutation motif sono corredati di tutta l’informazione relativa all’intero D-Loop. Importante è la doppia lettura dei polimorfismi, proveniente almeno da due ampliconi derivanti da due primers differenti. Ciò è necessario come “scudo” contro le “mutazioni fantasma” (*“phantom mutations”*) (Bandelt et al, 2002; Brandstätter et al, 2005).

c) Classificazione in aplogruppi (adeguata): come ribadito più volte, non sempre è sufficiente il sequenziamento della regione di controllo a tale scopo (Herrnstadt et

al, 2002; Coble et al, 2004), e ciò è apparso chiaro anche dai nostri dati; pertanto è consigliabile ottenere l'informazione anche dalla regione codificante, facendo riferimento alla gerarchia più aggiornata relativa agli aplogruppi. La ricerca degli SNPs, auspicabile in tal senso, potrebbe però portare a definirne un set troppo limitato per dare una definizione corretta o errato, in quanto alcune posizioni screenate sono troppo variabili (vedere le varie problematiche riscontrate anche nel presente studio: non ultimo l'ingente riscontro di mutazioni private o comunque di polimorfismi, rappresentati singolarmente). L'interpretazione di un dato set di SNP richiede dunque molta attenzione.

d) Controllo di qualità: sono state istituite delle linee guida ad opera dello SWGDAM (Scientific Working Group on DNA Analysis Methods) e dall'EMPOP in una fase successiva (Parson e Dür, 2007); questo riguarda le procedure operative nel sequenziamento dei campioni e nella loro interpretazione.

e) Analisi statistica: per essere significativa deve essere eseguita su database molto estesi e non essere limitata a singoli set di dati di medie dimensioni. In particolare, non ha senso produrre tali analisi per campioni piccoli (fino a 200 individui) (Pereira et al, 2004; 2009). Contrariamente ai marcatori nucleari classici, i siti dell'mtDNA non costituiscono marcatori indipendenti, ma sono correlati in base alla filogenia che ne è alla base, sebbene alcuni di essi siano soggetti pure a mutazioni ricorrenti. Pertanto, le tavole dei polimorfismi che ne riportano il numero in corrispondenza di una data posizione, hanno poco significato nel confronto dei DNA mitocondriali tra i diversi set di dati riferiti a diverse

popolazioni. La cosa migliore è invece fare riferimento alla filogenia aggiornata (Bandelt et al, 2006; Parson et al, 2007).

f) Documentazione ed annotazione: anche per queste sono stati formulati degli appositi criteri circa i formati elettronici per depositare le sequenze nei database correnti (GeneBank ed EMPOP) (Parson e Dür, 2007).

Capitolo 5

CONCLUSIONI E PROSPETTIVE

5.1 Conclusioni derivanti dal presente lavoro…

La Tabella IV.20 riassume le variazioni polimorfiche osservate nel DNA mitocondriale di ogni individuo.

La variabilità tra i 40 individui esaminati è elevata ed i siti polimorfici sono concentrati nella regione controllo o D-Loop (Sigurðardòttir et al, 2000). All'interno della regione codificante, il minor numero di variazioni è stato osservato nei geni dei tRNA e degli rRNA mitocondriali in accordo con gli studi sugli effetti patologici associati ad alterazioni di questi elementi genetici altamente conservati **(**McFarland et al, 2004; Wilson et al, 2004**).** Una variabilità relativamente maggiore è stata osservata nei geni codificanti le subunità enzimatiche dei complessi della catena respiratoria. Sono state individuate 8 mutazioni nuove, mai descritte in letteratura, di cui una con sospetta implicazione patologica. Le variazioni individuate come possibili siti di SNP sono coincidenti con quelle già descritte in letteratura ed inserite nell'albero filogenetico per l'individuazione degli aplogruppi. Quattro variazioni, apparentemente private, sono state viste essere condivise da soggetti accomunati da comune ascendente in via materna: su di esse, però, al momento, non è possibile fare alcuna previsione a livello di utilizzo come SNP, avendo solo riscontri scarsi ed eterogenei in Gene Bank. La variabilità del DNA

mitocondriale ha reso difficoltoso l'inquadramento in aplogruppi per tre soggetti (definiti "intermedi") ma nel contempo ha confermato, nel nostro campione, la presenza di meccanismi evolutivi specifici (retromutazioni), a livello di determinati siti, in altri 4 soggetti.

In HVIII sono state descritte una nuova eteroplasmia di lunghezza ed una nuova delezione, entrambe nello stesso soggetto e vicinissime come localizzazione, nonché una nuova inserzione nel DNA codificante di un altro soggetto. Sono stati inoltre descritti due casi di eteroplasmia di sequenza, noti solo come transizioni semplici omoplasmiche. Ulteriori studi consentiranno di chiarire i ruoli di queste tipologie di variazioni.

La classificazione in aplogruppi delle sequenze generate dimostra che l'aplogruppo H è il più frequente nella popolazione italiana, indipendentemente dalla suddivisione in aree geografiche: Nord, Centro, Sud e Isole. Tuttavia, la frequenza di aplogruppi rari (T,W,U) aumenta nell'Italia peninsulare. Un DNA mitocondriale ascrivibile all'aplogruppo L3d1b1, aplogruppo di origine africana e osservato molto raramente in Europa (Chen et al, 1995) è stato individuato fra gli individui provenienti dal Sud dalla penisola.

Infine è stato individuato un caso problematico per le varie tipologie di variazioni e per le implicazioni analitico-forensi che le medesime comportano.

Questa tesi è uno studio preliminare di un progetto più vasto il cui obiettivo è la determinazione della diversità del DNA mitocondriale nella popolazione italiana. I dati ottenuti saranno utilizzati per determinare i criteri di estensione del campionamento. La scelta di individui provenienti da zone chiave della penisola dovrebbe consentire

un’analisi statistica e molecolare completa della variabilità del DNA mitocondriale nel nostro Paese. L’accumulo di informazioni in tal senso potrebbe contribuire alla creazione/implementazione di un data base mitocondriale, (ricordiamo che diversi sono quelli esistenti, vedere Introduzione), mediante sequenze complete, che rispondano a procedure operative e criteri di qualità standard, utile a poter discriminare tra individui, in maniera analoga a quanto già avviene a livello del DNA nucleare e dei polimorfismi relativi al cromosoma Y.

Un approccio similare a quello descritto è parimenti auspicabile per altri gruppi di popolazione rilevanti da un punto di vista forense (ad es. Afro-Americani, Ispanici, Africani e così via, dato anche l’alto flusso di immigrazione verso il Nostro Paese), e possibilmente di sottopopolazioni (Slavi, ad esempio).

5.2 …Prospettive di maggiore portata

E’ stato dimostrato che i dati ricavabili dalla regione codificante possono essere risolutivi in un’ampia serie di casi: a) risoluzione di casi nei quali le famiglie di riferimento condividevano la stessa regione di controllo (Sturk et al, 2008; Just et al, 2009); b) selezione e ricomposizione di mescolanze di resti (Just et al, 2009); c) potenziato supporto statistico laddove i campioni di riferimento per un’eventuale esclusione non siano disponibili (Irwin et al, 2007); d) assegnazione dell’aplogruppo per un rapido screening di campioni provenienti da casi forensi (Brandstätter et al, 2003; Parson et al, 2007; Álvarez-Iglesias et al, 2007); e) la determinazione dell’ascendenza

materna bio-geografica come mezzo investigativo (Nelson et al, 2007; Köhnemann e Pfeiffer, 2010). Inoltre vi sono stati casi con quantità di materiale da testare estremamente limitante, per i quali si è scelto di esaminare, *in primis*, l'intera regione codificante di contestuali campioni di riferimento, presenti in quantità adeguate, ali fine di individuare la linea dell'aplogruppo; ottenuto questo, si è proceduto alla analisi di quei siti, spcifici dell'aplogruppo, direttamente sui campioni test (Just et al, 2010). In generale il sequenziamento di estese porzioni del genoma mitocondriale è poco pratico, laborioso e costoso, considerando anche il fatto che si ricorre a tale strumento in presenza di campioni scarsi o difficili. Questo è il motivo per il quale ci si è orientati verso la ricerca e la tipizzazione degli SNPs, mediante protocolli "home made" (Allen e Andreasson, 2004; Vallone et al, 2004; Salas et al, 2005) o mediante l'utilizzazione di kit commerciali (es. SNAPShot della Applied Biosystenms). Anche il presente lavoro ha cercato di dare un contributo in tal senso, provando ad identificarne ulteriori nuovi da inserire in questi protocolli. Questi sistemi, tuttavia, non sono esenti da limiti, sia per via della quantità di templato richiesta, sia per altri effetti correlati agli standard di qualità dei dati richiesti per le applicazioni in ambito forense (Vallone et al, 2007; Just et al, 2008). Per sopperire a tali difficoltà, sta prendendo piede una serie di tecnologie nuove mirate ad un agevole accesso all'informazione relativa all'intero genoma mitocondriale, e che rappresentano il futuro per tale tipo di indagine. Tra queste vi è una versione, per la regione codificante, dei saggi oligonucleotidici sequenza-specifici (Calloway et al, 2009), reazioni multiplex per la regione codificante da eseguire mediante la spettrometria di massa (Hall et al, 2005; Oberacher et al, 2006; Hall et al, 2009; Cerezo et al, 2009) e le cosiddette

tecnologie di Sequenziamento di Prossima Generazione (NGS). Il sequenziamento parallelo in forma massiva, consentito dalle NGS, sta rivoluzionando la produzione di dati genetici ed in un futuro non lontano consentirà il completo sequenziamento anche dei campioni degradati, con un rapporto temporale e di costi-benefici ottimali (Rogaev et al, 2009; Brenig et al, 2009; Krause et al, 2010). In prospettiva, quindi, l'applicazione dei dati relativi alla regione codificante del mtDNA nella routine dei casi forensi sarà dettata molto meno dalla quantità di materiale disponibile e/o dagli sforzi richiesti per ottenere i dati e molto più dalla disponibilità di estesi database di popolazione di alta qualità, relativi all'intero genoma mitocondriale, che può essere utilizzato per determinare la rarità degli aplotipi mitocondriali.

L'assenza di database con queste caratteristiche di fatto limita la comprensione dell'ulteriore valore discriminante che i dati relativi alla regione codificante possono fornire da individui presi a caso nella popolazione. Nel presente studio, per quanto effettuato su un esiguo campione, diverse sono state le difficoltà interpretative. Attualmente solo Gene Bank è l'unica depositaria di genomi mitocondriali completi che viene regolarmente aggiornata: prova ne è che nel presente studio, per avere informazioni complete sui polimorfismi identificati, si è dovuto ricorrere a più di un data base, e comunque senza mai prescindere da Gene Bank. Pur tuttavia la maggior parte delle sequenze presenti non sono state prodotte da individui presi a caso, o non imparentati che siano rappresentativi di particolari gruppi di popolazione. Inoltre, dal momento che molte non sono state sottoposte a controllo di qualità, non sono escludibili errori che possono alterare le stime di parametri statistici, in particolare le frequenze degli aplogruppi.

Anche i criteri di ricerca in Gene Bank, che non prevedono specifiche popolazioni di riferimento, inclusione/esclusione dei tratti poli-C, e range di sequenza predefiniti, sono limitanti e difficili nell'interfaccia BLAST. Perfino nuovi software che in qualche modo agevolano l'accesso ed il trattamento dei dati di sequenza mitocondriale presenti in Gene Bank, ad es. MitoVariome, sono limitati nell'allineamento laddove presenti le eteroplasmie di lunghezza (Lee et al, 2009; Bandelt et al, 2008). Gli sforzi attuali sono ancora diretti al miglioramento ed all'ampliamento dei set di dati forensi, accessibili a tutti e relativi alla regione di controllo, basti pensare al database EMPOP (http://www.empop.org) (Parson e Dür, 2007), che attualmente costituisce il riferimento principale a livello forense; pur tuttavia altrettanti sforzi devono a breve essere compiuti nella direzione della regione codificante, per stare al passo con le tecniche NGS. Ecco perché, in parallelo, è necessario sviluppare dati di popolazione di riferimento, relativi all'intero genoma mitocondriale, adatti per i confronti forensi e che aderiscano agli stessi standard di qualità dei dati già stabiliti per i database di popolazioni di riferimento relativamente alla regione di controllo (Parson et al, 2007; Irwin et al, 2007; Carracedo et al, 2010).

Il problema cruciale relativo alla regione codificante riguarda l'associazione con patologie. Sebbene le varianti patologiche, per la loro stessa natura, tendano ad essere selezionate negativamente, esse possono anche insorgere spontaneamente di volta in volta (e quindi quasi sempre le si ritrova in condizione eteroplasmica) e manifestarsi direttamente se quantitativamente rilevanti. Dato il ruolo focale dela funzione dell'mtDNA nelle funzioni basilari fisiologiche, il sequenziamento dei geni o di

frammenti di geni, al fine di aumentare la discriminazione a livello forense non può evitare di incontrare una variante che ora o forse in futuro, può essere associata ad una malattia o ad una condizione clinicamente rilevante (e nel presente studio ne sono state individuate diverse!). Ad esempio, una singola transizione alla posizione nucleotidica 14459, nel gene della NADH deidrogenasi 6 (ND6), che risulta in una variazione aminoacidica da alanina a valina, può provocare la cecità ad insorgenza improvvisa, denominata "Neropatia Ottica Ereditaria di Leber (LHON), come conseguenza della morte del nervo ottico (Wallace, 1999). L'età media di insorgenza è di 27.6 anni, con range effettivo tra 8 e 60. Ulteriori mutazioni dell'mtDNA associate a malattie genetiche ad insorgenza tardiva, collegate a mutazioni puntiformi che interessano i geni sull'mtDNA, includono il morbo di Parkinson (Brown et al, 1996), la malattia di Alzheimer (Lin et al, 2002) ed il diabete (Cavelier et al, 2002).

L'approccio inizialmente suggerito di selezionare solo i siti con mutazioni sinonime, paragonabili – per l'assenza di implicazioni patologiche – a quelli del D-loop (Coble et al, 2006), si scontra, oggi, con i diversi reports, sia pure con riserve (Bandelt et al, 2006; 2008; 2009), su Mitomap, di alcuni di questi siti in associazione a patologie. Il progresso delle conoscenze scientifiche può portare all'identificazione di nuove mutazioni patogeniche (Maragh et al, 2008), magari anche di quelle attualmente non ritenute tali. In linea di massima, tuttavia, l'approccio suddetto viene seguito come step iniziale nella maggior parte dei laboratori, anche se spesso deve fare i conti con l'assenza di database di popolazione ottimali per stabilire il peso di un dato polimorfismo della regione codificante. Tali database si rendono quindi estremamente indispensabili per il

lavoro dei vari laboratori, forensi o di ricerca o diagnostici, nonché per la formulazione di linee guida, anche a livello legislativo, per l'utilizzo dei dati relativi alla regione codificante. La produzione di siffatti set di dati non è così immediata, dal momento che le tecnologie attualmente in uso sono basate sul metodo di Sanger e, per quanto le nuove, quali la spettrometrias di massa, siano preferibili per i costi più bassi e la migliore resa, c'è da mettere in conto che i dati così prodotti saranno specifici per tale applicazione. Di conseguenza, e fintanto che le tecnologie NGS non vengono ottimizzate ed impiegate nei diversi laboratori, lo sforzo a breve termine sarà ancora orientato sulle tecnologie ed i protocolli attualmente in uso, fermo restando l'obiettivo di generare dati di alta qualità (Coble et al, 2004; Fendt et al, 2009). Gli sforzi in termini di tempo, fatica, fondi e risorse finanziarie sono ingenti, ma il ritorno, a lungo termine, sarà un insieme di interi genomi mitocondriali, di alta qualità, che si presenterà come un valido riferimento, per la comunità forense, per l'analisi del DNA mitocondriale e nel contempo rappresenterà una risorsa affidabile per l'ulteriore caratterizzazione della genetica di popolazione e dell'evoluzione molecolare, dal momento che queste sono strettamente correlate all'interpretazione del dato: distribuzione degli aplogruppi e tassi di sostituzione, solo per citare gli aspetti più importanti. Non è fuori luogo ricordare, a tal proposito, che da un lato i siti rari, lenti, hanno un alto potere discriminante, potendosi considerare come "aghi nel pagliaio", specifici per particolari tipi comuni, come, d'altro canto, i cosiddetti "hot spots" universali, ad evoluzione rapida(16519 e 152, tra i più rappresentativi) (Meyer et al, 1999; Allard et al, 2002; Malyarchuck et al, 2002), sarebbero utili non solo per risolvere tipi comuni fortemente correlati, ma anche tipi comuni di qualsiasi gruppo o

razza, in base all’estensione dell’eterogeneità del tasso evolutivo nella regione codificante (Berglund et al, 2009).

In parallelo sarà possibile migliorare sensibilmente le linee guida per l’interpretazione del mtDNA in ambito forense e massimizzare l’applicabilità dell’indagine sul mtDNA.

B I B L I O G R A F I A

Abu-Amero K.K. & Bosley T.M. (2006) **Mitochondrial abnormalities in patients with LHON-like optic neuropathies;** *Investigative Ophthalmology and Visual Science 47(10): 4211-20.*

Accetturo M., SantamariaM., Lascaro D., Rubino F., Achilli A., Torroni A., Tommaseo-Ponzetta M. and Attimonelli M. (2006) **Human mtDNA Site-Specific Variability Values Can Act as Haplogroup Markers**; *Hum Mut 27(9):965-74.*

Achilli A., Rengo C., Magri C., Battaglia V., Olivieri A., Scozzari R., Cruciani F., Zeviani M., Briem E., Carelli V. et al. (2004) **The molecular dissection of mtDNA Haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool;** *Am.J.Hum.Genet. 75:910-18*

Achilli A., Rengo C., Battaglia V., Pala M., Olivieri A., Fornarino S., Magri C., Scozzari R., Babudri N., Santachiara-Benerecetti A.S. et al (2005) **Saami and Berbers—an unexpected mitochondrial DNA link;** *Am J Hum Genet 76(5):883-6.*

Achilli A., Olivieri A., Pala M., Metspalu E., Fornarino S., Battaglia V., Accetturo M., Kutuev I., Khusnutdinova E., Pennarun E. et al. (2007) **Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans;** *Am J Hum Genet. 80:759-68.*

Achilli A., Perego U.A., Bravi C.M., Coble M.D., Kong Q.P., Woodward S.R., Salas A., Torroni A., Bandelt H.J. (2008) **The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies;** *PLoS ONE 3(3):e1764.*

Adachi J. & Hasegawa M. (1995) **Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites;** *J. Mol. Evol. 40:622-8.*

Ahari S.E., Houshmand M., Panahi M.S., Kasraie S., Moin M., Bahar M.A. (2007) **Investigation on mitochondrial tRNA(Leu/Lys), NDI and ATPase 6/8 in Iranian multiple sclerosis patients;** *Cell Mol Neurobiol. 27(6):695-700.*

Allard M.W., Miller K., Wilson M., Monson K. and Budowle B. (2002) **Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences**. Scientific Working Group on DNA Analysis Methods; *J. Forensic Sci. 47: 1215-23.*

Allen M. & Andreasson H. (2004) **Mitochondrial d-loop and coding sequence analysis using pyrosequencing**; *Methods Mol. Biol. 297:179–96.*

Álvarez-Iglesias V., Jaime J.C., Carracedo Á, Salas A. (2007) **Coding region mitochondrial coding region SNPs: Targeting East Asian and Native American haplogroups**; Forensic Sci. Int. Genet. 1: 44–55.

Álvarez-Iglesias V., Mosquera-Miguel A., Cerezo M., Quintáns B., Zarrabeitia M.T., Cuscó I., Lareu M.V., García O., Pérez-Jurado L., Carracedo A., Salas A. (2009) **New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0;** *PLoS ONE 4(4):e5112.*

Anderson S., Bankler A.T., Barrel B.G., de Bruijn M.H.L., Cuolson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F. et al. (1981) **Sequence and organization of the human mitochondrial genome;** *Nature 290:457-64.*

Andreasson H., Asp A., Alderborn A., Gyllensten U. and Allen M. (2002) **Mitochondrial sequence analysis for forensic identification using pyrosequencing technology;** *Biotechniques 32: 124-33.*

Andrews R.M., Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M. and Howell N. (1999) **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA;** *Nature Genetics 23:14.*

Aquadro C.F. & Greenberg, B.D. (1983) **Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals;** *Genetics 103: 287-312.*

Aris-Brosou S. & Excoffier L. (1996) **The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism;** *Mol. Biol. Evol. 13: 494-504.*

Asari M., Tan Y., Watanabe S., Shimizu K., Shiono H. (2007): **Effect of length variations at nucleotide positions 303-315 in human mitochondrial DNA on transcription termination**; *Biochem Biophys Res Commun. 361(3):641-4.*

Babalini C., Martinez-Labarga C., Tolk H.V., Kivisild T., Giampaolo R., Tarsi T., Contini I., Barac L., Janicijevic B., Martinovic Klaric I. et al. (2005) **The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view;** *Eur. J. Hum. Genet. 13:902-12.*

Ballard J.W.O. & Whitlock M.C. (2004) **The incomplete natural history of mitochondria;** *Molecular Ecology 13:729-44*

Bandelt H.J., Quintana-Murci L., Salas A. and Macaulay V. (2002) **The fingerprint of phantom mutations in mitochondrial DNA data;** *Am. J. Hum. Genet. 71: 1150-60.*

Bandelt HJ, Kong QP, Parson W, Salas A (2005): **More evidence for non-maternal inheritance of mitochondrial DNA?** *J.M.G. 42(12):957-60.*

Bandelt H.J., Salas A., Bravi C.M. (2006) **What is a 'novel' mtDNA mutation—and does 'novelty' really matter?** *J. Hum. Genet. 51:1073–82.*

Bandelt H.J. & Parson W. (2008) **Consistent treatment of length variants in the human mtDNA control region: a reappraisal;** *Int. J. Legal Med. 122:11–12.*

Bandelt H.J., Salas A., Taylor R.W. and Yao Y.G. (2008) **Exaggerated status of ‘‘novel’’ and ‘‘pathogenic’’ mtDNA sequence variants due to inadequate database searches**; *Hum. Mutat. 30, 191–6.*

Bandelt H.J., Yao Y.G., Salas A. (2008) **The search of ‘novel’ mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor**; *Int. J. Cardiol. 126:439–42.*

Bandelt H.J., Salas A., Taylor R.W., Yao Y.G. (2009) **Exaggerated status of ‘‘novel’’ and ‘‘pathogenic’’ mtDNA sequence variants due to inadequate database searches**; *Hum. Mutat. 30:191–6.*

Behar D.M., Metspalu E., Kivisild T., Achilli A., Hadid Y., Tzur S., Pereira L., Amorim A., Quintana-Murci L., Majamaa K. et al (2006) **The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event;** *Am J Hum Genet 78(3):487-97.*

Behar D.M., Metspalu E., Kivisild T., Rosset S., Tzur S., Hadid Y., Yudkovsky G., Rosengarten D., Pereira L., Amorim A. et al (2008a) **Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora**; *PLoS ONE 3(4):e2062.*

Behar D.M., Villems R., Soodyall H., Blue-Smith J., Pereira L., Metspalu E., Scozzari R., Makkan H., Tzur S., Comas D. et al. (2008b) **The dawn of human matrilineal diversity;** *Am J Hum Genet 82(5):1130-40.*

Bendall K.E. & Sykes B.C. (1995): **Length heteroplasmy in the first hypervariable segment of the human mtDNA control region**; *Am J Hum Genet. 57(2):248-56.*

Berglund J, Pollard KS, Webster MT (2009): **Hotspots of Biased Nucleotide Substitutions in Human Genes**; *PLOS Biology 7:45-62.*

Bilal E., Rabadan R., Alexe G., Fuku N.,Ueno H.,Nishigaki Y., Fujita Y., Ito M., Arai Y., Hirose N. et al. (2008) **Mitochondrial DNA haplogroups D4a is a marker for extreme longevity in Japan**; *PLoS ONE. 3, e2421*.

Bini C., Ceccardi S., Luiselli D., Ferri G., Pelotti S., Colalongo C., Falconi M., Pappalardo G. (2003) **Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy)**; *Forensic Science International 135 (2003) 48–52.*

Biondi G., Martini F., Rickards O. e Rotilio G. (2006) **In carne e ossa**. *Editori Laterza*

Bisbing R.E. (1982). **The forensic identification and association of human hair**. in Forensic Science Handbook, R. Saferstein, ed. (New Jersey: Prentice Hall Regents), *pp. 184-221*.

Brandon M., Baldi P., Wallace D.C. (2006) **Mitochondrial mutations in cancer;** *Oncogene 25(34):4647-62* .

Brandstätter A., Parsons T.J., Parson W. (2003) **Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups;** *Int. J. Legal Med. 117 (2003) 291–8.*

Brandstätter A., Peterson C.T., Irwin J.A., Mpoke S., Koech D.K., Parson W. and Parsons T.J. (2004) **Mitochondrial DNA control region sequence from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of forensic database**; *Int.J.Legal.Med. 118:294-306.*

Brandstätter A., Sänger T., Lutz-Bonengel S., Parson W., Béraud-Colomb E., Wen B., Kong Q.P., Bravi C.M., Bandelt H.J. (2005) **Phantom mutation hotspots in human mitochondrial DNA;** *Electrophoresis 26(18):3414-29.*

Brandstätter A., Zimmermann B., Wagner J., Göbel T., Röck A.W., Salas A., Carracedo A., Parson W. (2008) **Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East**; *BMC Evol Biol 8:191.*

Brenig B., Beck J., Schütz E. (2009) Shotgun **metagenomics of biological stains using ultra-deep DNA sequencing;** *Forensic Sci. Int. Genet., doi:10.1016/j.fsigen.2009.10.001*.

Brown M.D., Torroni A., Shoffner J.M., Wallace D.C. (1992) **Mitochondrial tRNAThr mutations and lethal infantile mitochondrial myopathy;** *American Journal of Human Genetics 51(2): 446-7* .

Brown M.D., Shoffner J.M., Kim Y.L., Jun A.S., Graham B.H., Cabell M.F., Gurley D.S. and Wallace D.C. (1996) **Mitochondrial DNA sequence analysis of four Alzheimer's and Parkinson's disease patients**; *Am. J. Med. Genet. 61: 283-89.*

Brown M.D., Hosseini S.H., Torroni A., Bandelt H.J., Allen J.C., Schurr T.G., Scozzari R., Cruciani F. and Wallace D.C. (1998) **MtDNA haplogroup X: an ancient link between Europe/Western Asia and North America?;** *Am.J.Hum.Genet. 63:1852-61.*

Brown W.M., George M. Jr. and Wilson A.C. (1979) **Rapid evolution of animal mitochondrial DNA;** *Proc. Natl. Acad. Sci. USA 76: 1967-71.*

Brown W.M. (1980) **Polymorphism in mitochondrial DNA of human as revealed by restriction endonuclease analysis;** *PNAS 77:3605-9.*

Calloway C., Stuart S., Erlich H. (2009) **Development of a multiplex PCR and linear array probe assay targeting informative polymorphisms within the entire mitochondrial genome**. *http://www.ncjrs.gov/pdffiles1/nij/grants/228279.pdf.*

Cann R.L., Stoneking M. and Wilson A.C. (1987) **Mitochondrial DNA and human evolution;** *Nature 325:31-6.*

Carracedo Á., Butler J.M., Gusmaõ L., Parson W., Roewer L., Schneider P.M. (2010) **Publication of population data for forensic purposes**; *Forensic Science International: Genetics 4: 145–7.*

Carrozzo R., Wittig I., Santorelli F.M., Bertini E., Hofmann S., Brandt U., Schägger H. (2005) **Subcomplexes of human ATP synthase mark mitochondrial biosynthesis disorders;** *Annals of Neurology 59(2): 265 – 75.*

Cavalli-Sforza L.L. & Feldman L.W. (2003) **The application of molecular genetic approaches to the study of human evolution;** *Nature Genetic Supplement 33: 266-75.*

Cavalli-Sforza L.L., Menozzi P. and Piazza A. (2005) **Storia e geografia dei geni umani**. *Gli Adelphi.*

Cavelier L., Erikson I., Tammi M., Jalonen P., Lindholm E., Jazin E., Smith P., Luthman H. and Gyllensten U. (2002) **MtDNA mutations in maternally inherited diabetes: presence of the 3397 ND1 mutation previously associated with Alzheimer's and Parkinson's disease;** *J. Neuropathol. Exp. Neurol. 61: 634-9.*

Cerezo M., Černý V., Carracedo A., Salas A. (2009) **Applications of MALDI-TOF MS to largescale human mtDNA population-based studies**; *Electrophoresis 30:3665–73.*

Chen B., Sun D., Yang L., Zhang C., Yang A., Zhu Y., Zhao J., Chen Y., Guan M., Wang X. et al (2008**) Mitochondrial ND5 T12338C, tRNA(Cys) T5802C, and tRNA(Thr) G15927A variants may have a modifying role in the phenotypic manifestation of deafness-associated 12S rRNA A1555G mutation in three Han Chinese pedigrees;** *American Journal of Medical Genetics. Part A . 146A (10): 1248-58*

.

Chen X.J. & Butow R.A. (2005) **The organization and inheritance of the mitochondrial genome** *Nature Reviews Genetics 6:815-25.*

Chen Y.S., Torroni A., Excoffier L., Santachiara-Benerecetti A.S. and Wallace D.C. (1995) **Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups;** *Am. J. Hum.Genet. 57:133-49.*

Chinnery P.F., Elliott H.R., Patel S., Lambert C., Keers S.M., Durham S.E., McCarthy M.I., Hitman G.A., Hattersley A.T. and Walker M. (2005) **Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes;** *The Lancet 366:1650-51.*

Chomyn A. & Attardi G. (2003) **MtDNA mutations in aging apoptosis;** *BBRC 304:519-29.*

Clayton D.A. (1982) **Replication of animal mitochondrial DNA;** *Cell. 28:693-705.*

Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T., Irwin J.A., Parsons T.J. (2004) **Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians***; Int J Legal Med* 118:137–46.

Coble M.D., Vallone P.M., Just R.S., Diegoli T.M., Smith B.C., Parsons T.J. (2006) **Effective strategies for forensic analysis in the mitochondrial DNA coding region**; *Int. J.Legal Med. 120:27–32.*

Dato S., Passarino G., Rose G., Altomare K., Bellizzi D., Mari V., Feraco E., Franceschi C. and De Benedictis G. (2004) **Association of the mitochondrial DNA haplogroup J with longevity is population specific;** *Eur. J. Hum. Genet. 12:1080-2.*

De Benedictis G., Rose G., Carreiri G., DeLuca M., Falcone E., Passarino G., Bonafe M., Monti D., Baggio G., Bertolini S. et al. (1999) **Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans**; *FASEB J. 13, 1532–6.*

Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Dambueva I., Perkova M., Dorzhu C., Luzina F., Lee H.K., Vanecek T. et al. (2007) **Phylogeographic analysis of mitochondrial DNA in northern Asian populations;** *Am J Hum Genet 81(5):1025-41.*

Di Mauro S. (2004) **Mitochondrial diseases;** Biochimica *et Biophysica Acta 1658:80-8.*

Elson J.L., Andrews R.M., Chinnery P.F., Lightowlers R.N., Turnbull D.M. and Howell N. (2001) **Analysis of European mtDNAs for recombination;** *Am. J. Hum. Genet. 68:145-53.*

Elson J.L., Herrnstadt C., Preston G., Thal L., Morris C.M., Edwardson J.A., Beal M.F., Turnbull D.M. and Howell N. (2006) **Does the mitochondrial genome play a role in the etiology of Alzheimer's disease?** *Hum. Genet. 119, 241–54.*

Ewing B., Hillier L., Wendl M.C., Green P. (1998): **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**; *Genome Research 8(3):175-85.*

Ewing B. & Green P. (1998): **Base-calling of automated sequencer traces using phred. II. Error probabilities**; *Genome Research 8(3):186-94.*

Excoffier L.& Yang Z. (1999) **Substitution rate variation among sites in the mitochondrial hypervariable region I of humans and chimpanzees**; *Mol. Biol. Evol. 16: 1357-68.*

Feder J., Blech I., Ovadia O., Amar S., Wainstein J., Raz I., Dadon S., Arking D.E., Glaser B., Mishmar D. (2008) **Differences in mtDNA haplogroup distribution**

among 3 Jewish populations alter susceptibility to T2DM complications; *BMC Genomics 9:198.*

Fendt L., Zimmermann B., Daniaux M., Parson W. (2009) **Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences;** *BMC Genomics 10:139.*

Fernàndez-Silva P., Enriquez J.A. and Montoya J. (2006) **Replication and transcription of mammalian mitochondrial DNA** *Experimental Physiology 88:41-56.*

Finnilä S., Hassinen I.E., Ala-Kokko L. and Majamaa K (2000): **Phylogenetic Network of the mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the Complete Coding Region by Conformation-Sensitive Gel Electrophoresis;** *Am. J. Hum. Genet. 66:1017–26.*

Finnilä S., Lehtonen M. and Majamaa K. (2001) **Phylogenetic Network for European mtDNA;** *Am. J. Hum. Genet. 68:1475–84.*

Fish J., Raule N. and Attardi G. (2004) **Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis;** *Science 306:2098-101.*

Fraumene C., Belle E.M., Castri L., Sanna S., Mancosu G., Cosso M., Marras F., Barbujani G., Pirastu M. and Angius A. (2006) **High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates;** *Mol Biol Evol. 23:2101-11.*

Fu Y.X. (1994) **A Phylogenetic Estimator of Effective Population Size or Mutation Rate;** *Genetics 136: 685-92.*

Gabriel M.N., Huffine E.F., Ryan J.H., Holland M.M. and Parsons T.J. (2001) **Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy;** *J. Forensic Sci. 46: 247-53.*

Ghezzi D., Marelli C., Achilli A., Goldwurm S., Pezzoli G., Barone P., Pellecchia M.T., Stanzione P., Brusa L., Bentivoglio A.R. et al. (2005) **Mitochondrial DNA haplogroup K is associated with a lower risk of Parkinson's disease in Italians;** *Eur. J. Hum. Genet. 13:748-52.*

Giles R.E., Blanc H., Cann H.M. and Wallace D.C. (1980) **Maternal inheritance of human mitochondrial DNA;** *Proc. Natl. Acad. Sci. USA 77: 6715-19.*

González A.M., Garcia O., Larruga J.M., Cabrera V.M. (2006) **The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country;** *BMC Genomics 7:124.*

González A.M., Larruga J.M., Abu-Amero K.K., Shi Y., Pestano J., Cabrera V.M. (2007) **Mitochondrial lineage M1 traces an early human backflow to Africa**; *BMC Genomics 8:223.*

Gray M.W., Burger G. and Franz Lang B. (1999) **Mitochondrial Evolution;** *Science 283:1476-81.*

Grignani P., Peloso G., Achilli A., Turchi C., Tagliabracci A., Alù M., Beduschi G., Ricci U., GiuntiL, Robino C. et al (2006) **Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification**; *Int J Legal Med 120: 151–6.*

Hall T.A., Budowle B., Jiang Y., Blyn L., Eshoo M., Sannes-Lowery K. et al. (2005) **Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans;** *Anal. Biochem. 344:53–69.*

Hall T.A., Sannes-Lowery K.A., McCurdy L.D., Fisher C., Anderson T., Henthorne A.et al. (2009) **Base composition profiling of human mitochondrial DNA using polymerase chain reaction and direct automated electrospray ionization mass spectrometry**; *Anal. Chem. 81:7515–26.*

Hasegawa M., Di Rienzo A., Kocher T.D. and Wilson A.C. (1993) **Toward a more accurate time scale for the human mitochondrial DNA tree;** *J. Mol. Evol. 37: 347-54.*

Helm M., Brulè H., Friede D., Giegè R., Pütz D. and Florentz C. (2000) **Search for characteristic structural features of mammalian mitochondrial tRNAs;** *RNA 6:1356-79.*

Herrnstadt C., Elson J.L., Fahy E., Preston G., Turnbull D.M., Anderson C., Ghosh S.S., Olefsky J.M., Beal M.F., Davis R.E. and Howell N. (2002): **Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups;** *Am. J. Hum. Genet. 70:1152–71.*

Heyer E., Zietkiewicz E., Rochowski A., Yotova V., Puymirat J. and Labuda D. (2001) **Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees**; *Am. J. Hum. Genet. 69: 1113-26.*

Holland M.M., Fisher D.L., Mitchell L.G., Rodriquez W.C., Canik J.J., Merril C.R. and Weedn V.W. (1993) **Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War;** *J. Forensic Sci. 38: 542-53.*

Holland M.M. & Parsons T.J. (1999) **Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework;** *Forensic Science Review 11: 21-50.*

Horai S. & Hayasaka K. (1990) **Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA;** *Am. J. Hum. Genet. 46: 828-42.*

Horai S., Hayasaka K., Kondo R., Tsugane K. and Takahata N. (1995) **Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs;** *Proc. Natl. Acad. Sci. USA 92: 532-36.*

Howell N., Kubacka I. and Mackey D.A. (1996) **How rapidly does the human mitochondrial genome evolve?** *Am. J. Hum. Genet. 59: 501-9.*

Howell N., Smejkal C.B., Mackey D.A., Chinnery P.F., Turnbull D.M. and Herrnstadt C. (2003) **The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates;** *Am. J. Hum. Genet. 72:659-70.*

Howell N., Elson J.L., Turnbull D.M. and Herrnstadt C. (2004): **African Haplogroup L mtDNA sequences show violations of clock-like evolution;** *Mol. Biol. Evol. 21(10):1843–54.*

Ingman M., Kaessmann H., Pääbo S. and Gyllensten U. (2000) **Mitochondrial genome variation and the origin of modern humans;** *Nature 408:708-13.*

Ingman M. & Gyllensten U. (2001) **Analysis of the complete human mtDNA genome: methodology and inferences for human evolution;** *J. Hered. 92: 454-61.*

Ingman M. & Gyllensten U. (2006) **mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences**; *Nucleic Acids Res. 34 (Database issue), D749–D751.*

Ingman M. & Gyllensten U. (2007) **Rate variation between mitochondrial domains and adaptive evolution in humans**; *Hum Mol Genet 16(19):2281–7.*

Innan H & Nordborg M (2002): **Recombination or mutational hot spots in human mtDNA?** *Mol. Biol. Evol. 19(7):1122–27.*

Irwin J., Saunier J., Strouss K., Sturk K., Diegoli T., Just R., Coble M., Parson W., Parsons T. (2007) **Development and expansion of high quality control region databases to improve forensic mtDNA evidence interpretation;** *Forensic Sci. Int. Genet. 1:154–7.*

Irwin J.A., Edson S.M., Loreille O., Just R.S., Barritt S.M., Lee D.A. et al. (2007) **DNA identification of ‘‘Earthquake McGoon’’ 50 years post-mortem;** *J. Forensic Sci. 52:1115–8.*

Jakupciak J.P., Maragh S., Markowitz M.E., Greenberg A.K., O Hoque M., Maitra A., Barker P.E., Wagner P.D., Rom W.N., Srivastava S., Sidransky D. and O'Connell C.D. (2008): **Performance of mitochondrial DNA mutations detecting early stage cancer;** *BMC Cancer 8:285-96.*

Janssen G.M.C., Neu A., ‘t Hart L.M., van de Sande C.M.T. and Maassen J.A. (2006) **Novel mitochondrial DNA length variants and genetic instability in a family with diabetes and deafness** *Exp.Clin.Endocrinol.Diabetes 114:168-74.*

Jazin E., Soodyall H., Jalonen P., Lindholm E., Stoneking M. and Gyllensten U. (1998) **Mitochondrial mutation rate revisited: hot spots and polymorphism;** *Nat. Genet. 18: 109-10.*

Johnson M.J., Wallace D.C., Ferris S.D., Rattazzi M.C. and Cavalli-Sforza L.L. (1983) **Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns;** *J. Mol. Evol. 19: 255-71.*

Just R.S., Lehrmann A.M., Harris K.E., Coble M.D. (2008) **Comparison of the AB mitoSEQr resequencing sets to standard mtDNA sequencing protocols**, *in: Presented at the English Speaking Working Group Meeting of the International Society for Forensic Genetics, Sinaia, Romania, 2008.*

Just R.S., Leney M.D., Barritt S.M., Los C.W., Smith B.C., Holland T.D. et al. (2009) **The use of mitochondrial DNA single nucleotide polymorphisms to assist in the resolution of three challenging forensic cases;** *J. Forensic Sci. 54:887–91.*

Just R.S., Loreille O.M., Molto J.E., Merriwether D.A., Woodward S.R., Matheson C.et al. (2010) **Titanic's unknown child: the critical role of the mitochondrial DNA coding region in a re-identification effort;** *Forensic Sci. Int. Genet., doi:10.1016/j.fsigen.2010.01.012.*

Kaguni L.S. (2004**) DNA polymerase γ, the mitochondrial replicase;** *Annu. Rev. Biochem. 73:293-320.*

Kasraie S., Houshmand M., Banoei M.M., Ahari S.E., Panahi M.S., Shariati P., Bahar M., Moin M. (2008) **Investigation of tRNA(Leu/Lys) and ATPase 6 genes mutations in Huntington's disease;** *Cell Mol Neurobiol. 28(7):933-8.*

Ke Y., Su B., Song X., Lu D., Chen L., Li H., Qi C., Marzuki S., Deka R., Underhill P. et al. (2001) **African origin of modern human in East Asia: a tale of 12000 Y chromosomes;** *Science 292:1151-3.*

Kimura M. (1983) **The Neutral Theory of Molecular Evolution**. *Cambridge: Cambridge University Press.*

Kivisild T., Shen P., Wall D.P., Do B., Sung R., Davis K., Passarino G., Underhill P.A., Scharfe C., Torroni A. et al. (2006) **The role of selection in the evolution of human mitochondrial genomes;** *Genetics 172:373-87.*

Kogelnik A.M., Lott M.T. Brown M.D., Navathe S.B. and Wallace D.C. (1998) **MITOMAP: a human mitochondrial genome database--1998 update;** *Nucleic Acids Res. 26: 112-5. (http://www.mitomap.org/)*

Köhnemann S. & Pfeiffer H. (2010) **Application of mtDNA SNP analysis in forensic casework**; *Forensic Sci. Int. Genet. doi:10.1016/j.fsigen.2010.01.015.*

Krause J., Briggs A.W., Kircher M., Maricic T., Zwyns N., Derevianko A. et al. (2010) **A complete mtDNA genome of an early modern human from Kostenki, Russia**; *Curr.Biol. 20:231–6.*

Krings M., Stone A., Schmitz R.W., Krainitzki H., Stoneking M. and Pääbo S. (1997) **Neandertal DNA sequences and the origin of modern humans;** *Cell 90:19-30.*

Krings M., Capelli C., Tschentscher F., Geisert H., Meyer S., von Haeseler A., Grosschmidt K., Possnert G., Paunovic M. and Pääbo S. (2000) **A view of Neandertal genetic diversity;** *Nat. Genet. 26:144-6.*

Kroemer G. (2003) **Mitochondrial control of apoptosis: an introduction;** *BBRC 304:433-5.*

Lee H.C. &Wei Y.H. (2009): **Mitochondrial DNA instability and metabolic shifts in human cancers;** *Int. J. Mol. Sci. 10: 674-701.*

Lee M.S. & Levin B.C. (2002) **MitoAnalyzer, a computer program and interactive web site to determine the effects of single nucleotide polymorphisms and mutations in human mitochondrial DNA;** *Mitochondrion 1: 321-6. (http://www.cstl.nist.gov/biotech/strbase/mitoanalyzer.html)*

Lee Y.S., Kim W.Y., Ji M., Kim J.H., Bhak J. (2009) **MitoVariome: a variome database of human mitochondrial DNA**; *BMC Genomics S12.*

Levin B.C., Cheng H. and Reeder D.J. (1999) **A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection**; *Genomics 55: 135-46.*

Levin B.C., Holland K.A., Hancock D.K., Coble M., Parsons T.J., Kienker L.J., Williams D.W., Jones M. and Richie K.L. (2003) **Comparison of the Complete mtDNA Genome Sequences of Human Cell Lines - HL-60 and GM10742A - From Individuals With Pro-Myelocytic Leukemia and Leber Heredity Optic Neuropathy, Respectively, and the Inclusion of HL-60 in the NIST Human Mitochondrial DNA Standard Reference Material - SRM 2392-I;** *Mitochondrion 2: 387-400.*

Lin M.T., Simon D.K., Ahn C.H., Kim L.M. and Beal M.F. (2002) **High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain**; *Hum. Mol. Genet. 11: 133-45.*

Liu Y., Cortopassi G., Steingrimsdottir H., Waugh A.P., Beare D.M., Green M.H., Robinson D.R., and Cole J. (1997) **Correlated mutagenesis of bcl2 and hprt loci in blood lymphocytes;** *Environ. Mol. Mutagen. 29: 36-45.*

Logan I. (2005) **The Medical Implications of Complete Mitochondrial DNA Sequencing**; *Journal of Genetic Genealogy 1:40-53.*

Logan J. (2009) **A refined phylogeny for mtDNA haplogroup J**. *Journal of Genetic Genealogy 5(1):16-22.*

Loogväli E.L., Roostalu U., Malyarchuk B.A., Derenko M.V., Kivisild T., Metspalu E., Tambets K., Reidla M., Tolk H.V., Parik J. et al. (2004). **Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia**; *Mol Biol Evol 21(11):2012-21.*

Lu, J., Qian, Y., Li, Z., Yang, A., Zhu, Y., Li, R., Yang, L., Tang, X., Chen, B., Ding, Y. et al. (2010) **Mitochondrial haplotypes may modulate the phenotypic manifestation of the deafness-associated 12S rRNA 1555A>G mutation;** *Mitochondrion 10(1): 69-81* .

Lundstrom R., Tavare S. and Ward R.H. (1992) **Modeling the evolution of the human mitochondrial genome**; *Math. Biosci. 112: 319-35.*

Lutz-Bonengel S., Schmidt U., Schmitt T., and Pollak S. (2003) **Sequence polymorphisms within the human mitochondrial genes MTATP6, MTATP8, and MTND4;** *Int. J. Leg. Med. 117: 133-42.*

Lutz-Bonengel S., Sänger T., Pollak S., Szibor R. (2004): **Different methods to determine length heteroplasmy within the mitochondrial control region;** *Int J Legal Med. 118(5):274-81.*

Maca-Meyer N., Gonzàlez A.M., Larruga J.M., Flores C. and Cabrera V.M. (2001) **Major genomic mitochondrial lineages delineate early human expansions;** *BMC Genetics 2:13.*

Macaulay V., Richards M., Hickey E., Vega E., Cruciali F,. Guida V., Scozzari R., Bonne´-Tamir B., Sykes B. and Torroni A. (1999a) **The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs;** *Am. J. Hum. Genet. 63:232-49.*

McFarland R., Elson J.L., Taylor R.W., Howell N. and Turnbull M. (2004) **Assigning pathogenicity to mitochondrial tRNA mutations: when "definitely maybe" is not good enough;** *Trends in Genetics 20:591-6.*

Malyarchuk B., Grzybowski T., Derenko M., Perkova M., Vanecek T., Lazur J., Gomolcak P., Tsybovsky I. (2008a) **Mitochondrial DNA phylogeny in Eastern and Western Slavs;** *Mol Biol Evol 25(8):1651-8.*

Malyarchuk B.A., Derenko M., Perkova M., Grzybowski T., Vanecek T., Lazur J. (2008b) **Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs;** *Eur J Hum Genet 16(9):1091-6.*

Malyarchuk B., Derenko M., Grzybowski T., Perkova M., Rogalla U., Vanecek T., Tsybovsky I. (2010a.) **The peopling of Europe from the mitochondrial haplogroup U5 perspective;** *PLoS One 5(4):e10285.*

Mancuso M., Filosto M., Mootha V.K., Rocchi A., Pistolesi S., Murri L., Di Mauro S. and Siciliano G. (2004) **A novel mitochondrial tRNA^{Phe} mutation causes MERRF sindrome;** *Neurology 62:2119-21.*

Maragh S., Jakupciak J.P., Wagner P.D., Rom W.N., Sidransky D., Srivastava S.and O'Connell C.D. (2008) **Multiple strand displacement amplification of mitochondrial DNA from clinical samples;** *BMC Medical Genetics 9:7.*

Marcuello A., Martínez-Redondo D., Dahmani Y., Casajús J.A., Ruiz-Pesini E., Montoya J., López-Pérez M.J., Díez-Sánchez C. (2009) **Human mitochondrial variants influence on oxygen consumption**; *Mitochondrion 9:27–30.*

Martínez-Redondo D., Marcuello A., Casajús J.A., Ara I., Dahmani Y, Montoya J., Ruiz-Pesini E., López-Pérez M.J., Díez-Sánchez C. (2010) **Human mitochondrial haplogroup H: The highest VO2max consumer – Is it a paradox?** *Mitochondrion 10:102–7.*

Mayr-Wohlfart U., Paulus C., Henneberg A., Rodel G. (1996) **Mitochondrial DNA mutations in multiple sclerosis patients with severe optic involvement;** *Acta Neurologica Scandinavica 94(3): 167-71* .

Mehta A.B., Vulliamy T., Gordon-Smith E.C. and Luzzatto L. (1989) **A new genetic polymorphism in the 16S ribosomal RNA gene of human mitochondrial DNA;** *Ann. Hum. Genet. 53: 303-10.*

Metspalu M., Kivisild T., Bandelt H.J., Richards M., Villems R. (2006). **The pioneer settlement of modern humans in Asia**. *In: Bandelt HJ, Macaulay V, Richards M, editors. Human mitochondrial DNA and the Evolution of Homo sapiens. Berlin: Springer-Verlag. p 181–99.*

Meyer S., Weiss G. and von Haeseler A. (1999) **Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA;** *Genetics 152: 1103-10.*

Meyer S. & Von Haeseler A. (2003) **Identifying site-specific substitution rates;** *Mol. Biol. Evol. 20: 182-9.*

Michaels G.S., Hauswirth W.W. and Laipis P.J. (1982) **Mitochondrial DNA copy number in bovine oocytes and somatic cells**; *Developmental Biology (Orlando) 94: 246-51.*

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A. and Wallace

D.C. (2003) **Natural selection shaped regional mtDNA variation in humans;** *Proc. Natl. Acad. Sci. USA 100: 171-6.*

Mitchell A.L., Elson J.L., Howell N., Taylor R.W., Turnbull D.M. (2006) **Sequence variation In mitochondrial complex I genes: mutation or polymorphism?;** *Journ Med Genet 43(2): 175-9* .

Mitchell P. (1961) **Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism;** *Nature 191:144-8.*

Moilanen J.S. & Majamaa K. (2003) **Phylogenetic network and physiochemical properties of nonsynonymous mutations in the protein coding genes of human mitochondrial DNA**; *Mol. Biol. Evol. 20, 1195–210.*

Monson K.L., Miller K.W.P., Wilson M.R., Di Zinno J.A. and Budowle B. (2002) **The mtDNA population database: an integrated software and database resource for forensic comparison;** *Forensic Sci. Comm., Vol 4, #2. (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm)*

Morelli L., Grosso M.G., Vona G., Varesi L., Torroni A. and Francalacci P. (2000) **Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia**; *Hum Biol. 72(4):585-95.*

Mostafaie N., Rossmanith W., Hombauer,H., Dechat T., Raffelsberger T., Bauer K., Worofka B., Kittl E., Hofmann J., Hejtman M. et al (2004) **Mitochondrial genotype**

and risk for Alzheimer's disease: cross-sectional data from the Vienna-Transdanube-Aging "VITA" study; *Journal of Neural Transmission 111 (9): 1155-1165* .

Nelson D.L. & Cox M.M. (2002) **I principi di biochimica di Lehninger**. *Zanichelli*

Nelson T.M., Just R.S., Loreille O., Schanfield M.S., Podini D. (2007) **Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing**; *Croatian Med. J. 48:460–72.*

Nishimura Y., Yoshinari T., Naruse K., Yamada T., Sumi K., Mitani H., Higashiyama T. and Kuroiwa T. (2006) **Active digestion of sperm mitochondrial DNA in single living sperm revealed by optical tweezers;** *PNAS 103:1382-7.*

Oberacher H., Niederstätter H., Pitterl F., Parson W. (2006) **Profiling 627 mitochondrial nucleotides via the analysis of a 23-plex polymerase chain reaction by liquid chromatography–electrospray ionization time-of-flight mass spectrometry**; *Anal.Chem. 78:7816–27.*

Olivieri A., Achilli A., Pala M., Battaglia V., Fornarino S., Al-Zahery N., Scozzari R., Cruciani F., Behar D.M., Dugoujon J.M. et al. (2006) **The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa;** *Science 314:1767-70.*

Ovchinnikov I.V., Gotherstrom A., Romanova G.P., Kharitonov V.M., Liden K. and Goodwin W. (2000) **Molecular analysis of Neanderthal DNA from the northern Caucasus;** *Nature 404:490-3.*

Paabo S. (1996) **Mutational hot spots in the mitochondrial microcosm;** *Am. J. Hum. Genet. 59: 493-6.*

Palanichamy M.G., Sun C., Agrawal S., Bandelt H.J., Kong Q.P., Khan F., Wang C.Y., Chaudhuri T.K., Palla V. and Zhang Y.P. (2004) **Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia**; *Am. J. Hum. Genet. 75:966–78.*

Parson W, Brandstätter A, Alonso A, Brandt N, Brinkmann B, Carracedo A, Corach D, Froment O, Furac I, Grzybowski T et al (2004): **The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives**; *Forensic Sci Int. 139(2-3):215-26.*

Parson W., Brandstätter A., Niederstätter, Grubwieser P., Scheithauer R. (2007) **Unraveling the mystery of Nanga Parbat**; *Int. J. Legal Med. 121:309–310.*

Parson W. & Dür A. (2007) **EMPOP—A forensic mtDNA database**; *Forensic Science International: Genetics 1: 88–92.*

Parsons T.J., Muniec D.S., Sullivan K., Woodyatt N., Alliston-Greiner R., Wilson M.R., Berry D.L., Holland K.A., Weedn V.W., Gill P. and Holland M.M. (1997) **A high observed substitution rate in the human mitochondrial DNA control region;** *Nature Genetics 15: 363-8.*

Parsons T.J. & Holland M.M. (1998) **Mitochondrial mutation rate revisited: hot spots and polymorphism**; *Nat. Genet. 18: 110.*

Penny D., Steel M., Waddell P.J. and Hendy M.D. (1995) **Improved analyses of human mtDNA sequences support a recent African origin for Homo sapiens;** *Mol. Biol. Evol. 12: 863-82.*

Perego U.A., Achilli A., Angerhofer N., Accetturo M., Pala M., Olivieri A., Hooshiar Kashani B., Ritchie K.H., Scozzari R., Kong Q.P. et al. (2009) **Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups;** *Curr Biol 19(1):1-8.*

Pereira L., Cunha C., AmorimA. (2004) **Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database**; *Int. J. Legal Med. 118 (3):132–6.*

Pereira L., Gonc¸alves J., Franco-Duarte R., Silva J., Rocha T., Arnold C., Richards M. and Macaulay V. (2007) **No evidence for an mtDNA role in sperm motility: Data from complete sequencing of asthenozoospermic males;** *Mol. Biol. Evol. 24, 868–74.*

Pereira L., Freitas F., Fernandes V., Pereira J.B., Costa M.D., Costa S., Máximo V., Macaulay V., Rocha R. and Samuels D.C. (2009): **The Diversity Present in 5140 Human Mitochondrial Genomes**; *Am Journ Hum Genet 84:628–40.*

Pesole, G. & Saccone C. (2001) **A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences;** *Genetics 157: 859-65.*

Petros J.A., Baumann A.K., Ruiz-Pesini E., Amin M.B., Sun C.Q., Hall J., Lim S., Issa M.M., Flanders W.D., Hosseini S.H. et al (2005) **mtDNA mutations increase tumorigenicity in prostate cancer;** *PNAS 102(3): 719-24.*

Piazza A., Cappello N., Olivetti E. and Rendine S. (1988) **A genetic history of Italy;** *Ann. Hum. Genet. 52:203-13.*

Piko L. & Matsumoto L. (1976) Number **of mitochondria and some properties of mitochondrial DNA in the mouse egg;** *Developmental Biology (Orlando) 49: 1-10.*

Reidla M., Kivisild T., Metspalu E., Kaldma K., Tambets K., Tolk H.V., Parik J., Loogväli E.L., Derenko M., Malyarchuk B. et al. (2003) **Origin and diffusion of mtDNA haplogroup X;** *Am J Hum Genet 73(5):1178-90.*

Relethford J.H. (2008) **Genetic evidence and the modern human origins debate;** *Heredity, 100:555–63.*

Richards M.B., Macaulay V.A., Bandelt H.J. and Sykes B.C. (1998) **Phylogeography of mitochondrial DNA in western Europe;** *Ann. Hum. Genet. 62: 241-60.*

Richards M., Macaulay V., Hickey E., Vega E., Sykes B., Guida V., Rengo C., Sellitto D., Cruciani F., Kivisild T. et al. (2000) **Tracing European founder lineages in the Near Eastern mtDNA pool;** *Am. J. Hum. Genet. 67:1251-76.*

Richards M. & Macaulay V. (2001) **The mitochondrial gene tree comes of age;** *Am. J. Hum. Genet. 68:1315–20.*

Rieder M.R. Taylor S.L., Tobe V.O. and Nickerson D.A. (1998) **Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome;** *Nucleic Acids Research 26:967-73.*

Rogaev E.I., Grigorenko A.P., Faskhutdinova G., Kittler E.L.W., Moliaka Y.K. (2009) **Genotype analysis identifies the cause of the ‘‘royal disease’’**; *Science 326: 817.*

Rollins B., Martin M.V., Sequeira P.A., Moon E.A., Morgan L.Z., Watson S.J., Schatzberg A., Akil H., Myers R.M., Jones E.G. et al (2009) **Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder;** *PLoS One 4(3): e4913* .

Roostalu U., Kutuev I., Loogvali E.L., Metspalu E., Tambets K., Reidla M., Khusnutdinova E.K, Usanga E., Kivisild T. and Villems R. (2007) **Origin and expansion of Haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The Near Eastern and Caucasian prospective;** *Mol. Biol. Evol. 24:436-48.*

Ruiz-Pesini E., Mishmar D., Brandon M., Procaccio V. and Fallace D.C. (2004) **Effects of purifying and adaptive selection on regional variation in human mtDNA**; *Science 303:223-6.*

Ruiz-Pesini E. & Wallace D.C. (2006) **Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA** *Hum. Mutat. 27, 1072–81.*

Ruiz-Pesini E., Lott M.T., Procaccio V.,Poole J.C., Brandon M.C., Mishmar D., Yi C., Kreuziger J., Baldi P. and Wallace D.C. (2007) **An enhanced MITOMAP with a**

global mtDNA mutational phylogeny; *Nucleic Acids Res. 35 (Database issue), D823–D828.*

Saccone C., Gissi C., Lanave C., Larizza A., Pesole G. and Reyes A. (2000) **Evolution of the mitochondrial genetic system: an overview**; *Gene 261: 153-9.*

Safaei S., Houshmand M., Banoei M.M., Panahi M.S., Nafisi S., Parivar K., Rostami M., Shariati P. (2009) **Mitochondrial tRNALeu/Lys and ATPase 6/8 gene variations in spinocerebellar ataxias**; *Neurodegener Dis. 6(1-2):16-22.*

Saillard J., Magalhaes P.J., Schwartz M., Rosenberg T. and Norby S. (2000) **Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV;** *Hum. Biol. 72: 1065-8.*

Salas A., Quintáns B., Álvarez-Iglesias V. (2005) **SNaPshot typing of mitochondrial DNA coding region variants**, *in: A. Carracedo (Ed.), Forensic DNA Typing Protocols, Methods in Molecular Biology, Humana Press, Totowa, New Jersey, pp. 197–208.*

Scheffler I.E. (1999) **Mitochondria.** *New York: Wiley-Liss, Inc.*

Schmitz R.W., Serre D., Bonani G., Feine S., Hillgruber F., Krainitzki H., Paabo S. and Smith F.H. (2002) **The Neandertal type site revisited: interdisciplinary**

investigations of skeletal remains from the Neander Valley, Germany; *Proc. Natl. Acad. Sci. U S A. 99:13342-7.*

Schurr T.G., Ballinger S.W., Gan Y.Y., Hodge J.A., Merriwether D.A., Lawrence D.N., Knowler W.C., Weiss K.M. and Wallace, D.C. (1990) **Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages;** *Am. J. Hum. Genet. 46: 613-23.*

Schwartz M. & Vissing J. (2002) **Paternal inheritance of mitochondrial DNA;** *N Engl J Med 22:576-80.*

Sigurðardòttir S., Helgason A., Gulcher J., Stefansson K. and Donnely P. (2000) **The mutation rate in the human mtDNA control region;** *Am.J.Hum.Genet. 66:1599-609.*

Springer M.S. & Douzery E. (1996) **Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules;** *J. Mol. Evol. 43:357-73.*

Stewart J.E., Fisher C.L., Aagaard P.J., Wilson M.R., Isenberg A.R., Polanskey D., Pokorak E., Di Zinno J.A. and Budowle B. (2001) **Length variation in HV2 of the human mitochondrial DNA control region;** *J. Forensic Sci. 46: 862-70.*

Sturk K.A., Coble M.D., Barritt S.M., Parsons T.J., Just R.S. (2008) **The application of mtDNA SNPs to a forensic case;** *Forensic Sci. Int. Genet. Supplement Series 1:295–7.*

Suissa S., Wang Z., Poole J., Wittkopp S., Feder J., Shutt T.E., Wallace D.C., Shadel G.S., Mishmar D. (2009) **Ancient mtDNA Genetic Variants Modulate mtDNA Transcription and Replication;** *PLoS Genet 5(5): e1000474.*

Sutovsky P., Moreno R.D., Ramalho-Santos J., Dominko T., Simerly C. and Schatten G. (2000) **Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos;** *Biology of reproduction 63:582-90.*

Tamm E., Kivisild T., Reidla M., Metspalu M., Smith D.G., Mulligan C.J., Bravi C.M., Rickards O., Martinez-Labarga C., Khusnutdinova E.K. et al. (2007) **Beringian standstill and spread of Native American founders;** *PLoS ONE 2(9):e829.*

Tanaka N., Goto Y.I., Akanuma J., Kato M., Kinoshita T., Yamashita F., Tanaka M., Asada T. (2010) **Mitochondrial DNA variants in a Japanese population of patients with Alzheimer's disease;** *Mitochondrion 10(1):32-7* .

Tang S., Batra A., Zhang Y., Ebenroth E.S., Huang T. (2010) **Left ventricular noncompaction is associated with mutations in the mitochondrial genome** *Mitochondrion 10(4): 350-7* .

Taylor R.W. & Turnbull D.M. (2005) **Mitochondrial DNA mutations in human disease;** *Nature Rev. Genet. 6:389-402.*

Templeton AR (1992) **Human origins and analysis of mitochondrial DNA sequences;** *Science 255: 737.*

Thomas A.W., Edwards A., Sherratt E.J., Majid A., Gagg J. and Alcolado J.C. (1996) **Molecular scanning of candidate mitochondrial tRNA genes in type 2 (non-insulin dependent) diabetes mellitus;** *J.Med.Genet. 33:253-5.*

Thompson J.D., Higgins D.G. and Gibson T.J. (1994) **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice;** *Nucleic Acids Research 22:4673-80.*

Torroni A., Schurr T.G., Cabell M.F., Brown M.D., Neel J.V., Larsen M., Smith D.G., Vullo C.M. and Wallace D.C. (1993) **Asian affinities and continental radiation of the four founding Native American mtDNAs;** *Am J Hum Genet. 53:563-90.*

Torroni A., Lott M.T., Cabell M.F., Chen Y., Lavergne L. and Wallace D.C. (1994) **mtDNA and the Origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to recurrent somatic duplication in the D-loop region;** *Am. J. Hum. Genet. 55:760-76.*

Torroni A., Huopenen K., Francalacci P., Petrozzi M., Morelli L., Scozzari R., Obinu D., Savontaus M.L. and Wallace D. (1996) **Classification of European mtDNAs from an analysis of three European populations;** *Genetics 144:1835-50.*

Torroni A., Richards M., Macaulay V., Forster P., Villems R., Nørby S., Savontaus M.L., Huoponen K., Scozzari R. and Bandelt H.J. (2000) **mtDNA haplogroups and frequency patterns in Europe;** *Am.J.Hum.Genet. 66:1173-7.*

Torroni A., Bandelt H.J., Macaulay V., Richards M., Cruciani F., Rengo C., Martinez-Cabrera V., Villems R., Kivisild T., Metspalu E. et al. (2001) **A signal, from human mtDNA, of postglacial recolonization in Europe;** *Am.J.Hum.Genet. 69:844-52.*

Torroni A., Achilli A., Macaulay V., Richards M. and Bandelt H.J. (2006) **Harvesting the fruit of the human mtDNA tree;** *Trends in Genetics 22:339-45.*

Trifunovic A. (2006) **Mitochondrial DNA and ageing;** *Biochimica et Biophysica Acta.*

Vallone P.M., Just R.S., Coble M.D., Butler J.M., Parsons T.J. (2004) **A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome**; *Int J Legal Med 118 : 147–57.*

Vallone P.M., Jakupciak J.P., Coble M.D. (2007) **Forensic application of the Affymetrix human mitochondrial resequencing array**; *Forensic Sci. Int. Genet. 1:196–8.*

Vernesi C., Caramelli D., Dupanloup I., Bertorelle G., Lari M., Cappellini E., Moggi-Cecchi J., Chiarelli B., Castri L., Casoli A., Mallegni F., Lalueza-Fox C. and Barbujani G. **The Etruscans: a population-genetic study;** *Am. J. Hum. Genet. 74:694-704.*

Vigilant L., Stoneking M., Harpending H., Hawkes K. and Wilson A.C. (1991) **African populations and the evolution of human mitochondrial DNA;** *Science 253:1503-7.*

Wakeley J. (1993) **Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA**; *J. Mol. Evol. 37: 613-23.*

Wakeley J. (1994) **Substitution-rate variation among sites and the estimation of transition bias;** *Mol. Biol. Evol. 11: 436-42.*

Wallace D.C., Ye J.H., Neckelmann S.N., Singh G., Webster K.A. and Greenberg B.D. (1987) **Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations;** *Curr. Genet. 12:81-90.*

Wallace D.C. (1994): **Mitochondrial DNA sequence variation in human evolution and disease;** *PNAS 91:8739-46.*

Wallace D.C., Brown M.D. and Lott M.T. (1999) **Mitochondrial DNA variation in human evolution and disease;** *Gene 238: 211-30.*

Wallace D.C. (2005) **A mitochondrial paradigm of metabolic and degenerative dieases, aging, and cancer: a dawn for evolutionary medicine;** *Annu. Rev. Genet. 39:359-407.*

Wang, X., Lu, J., Zhu, Y., Yang, A., Yang, L., Li, R., Chen, B., Qian, Y., Tang, X., Wang, J., Zhang, X., Guan, M. X. (2008) **Mitochondrial tRNAThr G15927A mutation may modulate the phenotypic manifestation of ototoxic 12S rRNA A1555G mutation in four Chinese families;** *Pharmacogenetics and Genomics 18(12): 1059-70* .

Watson E., Forster P., Richards M. and Bandelt H.J. (1997) **Mitochondrial footprints of human expansions in Africa;** *Am. J. Hum. Genet. 61: 691-704.*

Wei Y.H., Wu S.B., Ma Y.S., Lee H.C. (2009) **Respiratory Function Decline and DNA Mutation in Mitochondria, Oxidative Stress and Altered Gene Expression during Aging;** *Chang Gung Med J Vol. 32 No. 2: 113-32.*

Wiesbauer M., Meierhofer D., Mayr J.A., Sperl W., Paulweber B., Kofler B (2006) **Multiplex primer extension analysis for rapid detection of major European mitochondrial haplogroups**; *Electrophoresis 27: 3864–8.*

Wilson A.C., Cann R.L., Carr S.M., George M., Gyllensten U.B., Helm-Bychowsky K.M., Higuchi R.G., Palumbi S.R., Prager E.M., Sage R.D. and Stoneking M. (1985) **Mitochondrial DNA and two prospectives on evolutionary genetics;** *Biol. J. Linn. Soc. 26:375-400.*

Wilson F.H., Hariri A., Farhi A., Zhao H., Falk Petersen K., Toka H.R., Nelson-Williams C., Raja K.M., Kashgarian M., Shulman G.I., Scheinman S.J. and Lifton R.P.

(2004) **A cluster of metabolic defects caused by mutation in a mitochondrial tRNA;** *Science 306:1190-4.*

Wilson M.R., Holland M.M., Stoneking M., Di Zinno J.A. and Budowle B. (1993) **Guidelines for the use of mitochondrial DNA sequencing in forensic science;** *Crime Laboratory Digest 20: 68-77.*

Yang Z. (1996) **Among-site rate variation and its impact on phylogenetic analyses;** *TREE 11: 367-72.*

Yao Y.G., Macauley V., Kivisild T., Zhang Y.P. and Bandelt H.J. (2003) **To trust or not to trust an idiosyncratic mitochondrial data set;** *Am. J. Hum. Genet. 72: 1346-9.*

Yoon, K. L., Aprille, J. R., Ernst, S. G. (1991) **Mitochondrial tRNAThr mutation in fatal infantile respiratory enzyme deficiency;** *Biochemical and Biophysical Research Communications 176(3): 1112-5* .

Zuker M. (2003) **Mfold web server for nucleic acid folding and hybridization prediction;** *Nucleic Acids Res. 31:3406–15.*