

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
Scienze Biomediche e Neuromotorie**

Ciclo XXIX

Settore Concorsuale di afferenza: 05/H1

Settore Scientifico disciplinare: BIO/16

**ANALISI GENETICO-FORENSE SU DNA MITOCONDRIALE
APPARTENENTE ALLA POPOLAZIONE MONGOLA**

Presentata da: Dott. Francesco Carano

Coordinatore Dottorato

Prof. Lucio Cocco

Relatore

Prof.ssa Mirella Falconi

Correlatore

Prof.ssa Susi Pelotti

Esame finale anno 2017

Indice

Introduzione	6
Capitolo 1: Il mitocondrio	
1.1 Struttura e funzione	9
1.2 Genoma mitocondriale	13
1.3 Ereditarietà	17
1.4 Eteroplasmie	19
Capitolo 2: Applicazione del DNA mitocondriale	
2.1 Il DNA mitocondriale nello studio di popolazione e la sua filogenesi	23
2.2 Il DNA mitocondriale in Genetica Forense, analisi ed applicazioni	27
2.3 Il DNA mitocondriale in Genetica Medica e le patologie associate	31
Capitolo 3: Analisi del DNA mitocondriale	
3.1 Il Polimorfismo a singolo nucleotide	34
3.2 Il sequenziamento di Sanger nell'analisi del DNA mitocondriale	35
3.3 Tecnologie di <i>Next Generation Sequencing</i> ed applicazioni	38
3.4 La sequenza di riferimento di Cambridge	39
Capitolo 4: Scopo dello Studio	41
Capitolo 5: Materiali e metodi	
5.1 Campioni, provenienza ed estrazione	43
5.2 Preparazione della libreria genomica	43
5.3 Sequenziamento con Ion PGM System™	47
5.4 Analisi ed interpretazione delle varianti tramite NGS	49
5.5 Sequenziamento di Sanger	50
5.6 Calcolo dei parametri genetico-forensi	53

Capitolo 6: Risultati

6.1 Analisi filogenetica e molecolare del genoma mitocondriale 54

6.2 Analisi del genoma mitocondriale con *Next Generation Sequencing* 56

Capitolo 7: Discussione 61

Capitolo 8: Conclusioni 66

Appendice 68

Bibliografia 91

Introduzione

Da quando l'intera molecola del genoma mitocondriale umano è stata sequenziata da Anderson *et al.* nel 1981 [1], l'analisi di sequenza nei successivi vent'anni è diventata un importante strumento per merito di due caratteristiche peculiari quali: l'elevato tasso di mutazione rispetto al DNA nucleare e l'assenza di ricombinazione che hanno reso possibile la sua applicazione nella genetica di popolazione ed antropologia molecolare. In questi settori, l'analisi delle mutazioni puntiformi presenti sulla regione non codificante di controllo (D-loop), trasmesse inalterate nel corso delle generazioni lungo la linea materna, ha infatti consentito lo studio della struttura genetica e dell'origine della popolazione umana, fornendo una definitiva prova in supporto al modello *Out of Africa* [2].

Allo stesso modo, le ridotte dimensioni del genoma (16569 bp) e la sua presenza in più copie rispetto al DNA nucleare, lo hanno reso un eccellente marcatore sia per lo studio del DNA antico nella paleogenetica [3] per cui è possibile ricostruire la storia degli eventi migratori sia in genetica forense ai fini di identificazione personale [4] nel caso in cui il DNA nucleare estratto dalla traccia biologica sia particolarmente degradato. Infine, non può essere omissa l'assoluta ruolo primario dell'analisi della sequenza del genoma mitocondriale in genetica medica per lo studio della grande varietà delle patologie associate al mitocondrio [5].

Diversi metodi analitici sono stati applicati nel corso degli anni per l'analisi del DNA mitocondriale. I primi studi evoluzionistici facevano uso di enzimi di restrizione (RFLP) specifici per i polimorfismi di lunghezza [6]. Attualmente il *gold standard* è rappresentato dal metodo automatizzato del sequenziamento di Sanger tramite elettroforesi capillare, che utilizza dideossinucleotidi marcati con fluorocromo per la rilevazione di eventuali polimorfismi a singolo

nucleotide (SNPs) che verranno evidenziati in seguito, attraverso il confronto con la sequenza di riferimento *revised Cambridge Reference Sequence* (rCRS).

A partire dal 2005, cominciò a svilupparsi un progressivo allontanamento dal metodo di Sanger in favore di nuove tecnologie di seconda generazione, capaci di offrire una maggiore resa analitica a costi ridotti e per questa ragione chiamati *Next Generation Sequencing* (NGS), che per la loro capacità di analizzare milioni di frammenti di DNA contemporaneamente vengono anche definiti come *Massive Parallel Sequencing* (MPS). Il principale vantaggio consiste nel fatto che l'NGS consente il sequenziamento dell'intero mitogenoma superando i limiti tecnici imposti dal sequenziamento di Sanger nell'affrontare l'analisi di regioni estese e rappresentando un vantaggio in ambito di una migliore resa analitica e costi ridotti. Queste caratteristiche si rivelano quindi estremamente importanti oltre che nel campo biomedico, anche nell'antropologia molecolare, poiché il sequenziamento dell'intero DNA mitocondriale permette una migliore discriminazione tra le linee filogenetiche materne, caratteristica di grande rilevanza anche nel campo della genetica forense, in quanto la sola analisi ristretta alle regioni ipervariabili I e II (HVI; HVII) nella regione di controllo (CR) offre una scarsa risoluzione delle numerose linee materne mitocondriali, portando così a possibili interpretazioni errate nel caso in cui due soggetti apparentemente simili per le regioni analizzate appartengano in realtà a due linee filogenetiche diverse.

In questa tesi di dottorato verrà presentato uno studio condotto in collaborazione con il laboratorio di Genetica Forense dell'Istituto di Medicina Legale dell'Università di Medicina di Innsbruck ed il Dipartimento di Chimica, Biologia e Biotecnologie dell'Università di Perugia, assieme alla partecipazione del laboratorio di genetica forense del Dipartimento di Scienze Mediche e Chirurgiche dell'Università di Bologna, che ha riguardato il sequenziamento dell'intero mitogenoma in un campione di 151 individui originari della Mongolia, utilizzando la recente tecnologia di *Next Generation Sequencing* assieme ad un pannello sperimentale per l'analisi genetica.

Nel seguente lavoro, verrà valutata la differente resa analitica tra Sanger ed NGS e saranno anche discussi gli aspetti legati all'interpretazione dei risultati in termini di risoluzione e potere discriminativo, ottenuto dall'analisi dell'intero genoma mitocondriale e dalla regione di controllo, normalmente utilizzata in genetica forense per compiere una precisa associazione statistica tra due profili.

I risultati, peraltro, potranno avere importanti riflessi anche in più vasti ambiti della ricerca biomedica in cui il metodo di Next Generation Sequencing sta progressivamente prendendo il sopravvento, rispetto al tradizionale metodo di sequenziamento.

Capitolo 1: Il mitocondrio

1.1 *Struttura e funzione*

I mitocondri sono organelli intracellulari presenti in tutte le cellule nucleate dei mammiferi, con un proprio DNA e ribosomi. L'ipotesi più accreditata riguardo alla loro origine consiste nell'ingresso accidentale di un alfa-protobatterio all'interno di una cellula eucariote avvenuta circa due miliardi di anni fa, diventandone così endosimbionte [7]. Non possiedono una vera morfologia predefinita, tuttavia negli epatociti e nei fibroblasti sono stati osservati mitocondri di forma allungata, dalle dimensioni di circa 3-4 μm di lunghezza e 1 μm di diametro. La struttura del è caratterizzata dalla presenza di due membrane concentriche, altamente specializzate e con differenti funzioni, formate da un doppio strato lipidico. È possibile quindi distinguere: una membrana esterna a contatto con il citosol, permeabile a molecole di massa non superiori ai 5000 Da e piccole proteine, sulla quale sono localizzate numerosi canali proteici transmembrana definiti *porine* assieme ad apparati impiegati sia per la comunicazione con altre strutture cellulari sia per il riconoscimento e l'importazione di proteine mitocondriali codificate dal nucleo [8]. Al di sotto della membrana esterna, si localizza un primo compartimento chiamato spazio inter-membrana delimitato dalla membrana interna, in cui si localizzano diverse proteine di trasporto che la rendono selettiva per piccole molecole. Inoltre, il *bilayer* lipidico di cui è composta contiene una elevata quantità di cardiolipina, un fosfolipide che possiede nella sua struttura quattro molecole di acidi grassi che conferiscono alla membrana la capacità di essere fortemente impermeabile agli ioni. A livello strutturale, la membrana interna forma delle estroflessioni di aspetto lamellare definite creste mitocondriali [9], che oltre ad aumentarne la superficie, si approfondano all'interno della *matrice*: un secondo compartimento delimitato dalla membrana stessa, contenente un elevato numero di enzimi e copie multiple del

genoma mitocondriale codificante per le proteine della membrana interna, assieme al relativo RNA necessario per la traduzione. La membrana e le creste mitocondriali sono collegate da sottili connessioni definite giunzioni delle creste, benchè queste strutture siano ubiquitarie, la loro morfologia è variabile a seconda della tipologia cellulare, Perkins *et al.* hanno osservato che nei mitocondri presenti all'interno dei neuroni, le giunzioni delle creste possiedono una sezione circolare uniforme dal diametro di circa 28nm [10], mentre negli epatociti di ratto, la sezione passa da circolare ad ellittica con un diametro di circa 20-50 nm [11]. Anche le creste mitocondriali possiedono una morfologia variabile: una struttura tubulare semplice, prevalente nei mitocondri degli epatociti [12;13], oppure a lamina appiattita originata dalla fusione di più creste tubulari di grandezza variabile e proporzionale alla quantità di giunzioni delle creste incorporate [14]. Ampie strutture lamellari sono tipicamente osservabili nel tessuto adiposo bruno ed in quello nervoso, disposte parallele tra loro ed orientate perpendicolari all'asse maggiore del mitocondrio. Il numero e la struttura delle creste mitocondriali e della membrana interna riflettono verosimilmente la risposta alla domanda di energia da parte della cellula dove quindi è più elevata la produzione di Adenosina Trifosfato (ATP) attraverso fosforilazione ossidativa, la cui parte iniziale è svolta dalla catena respiratoria composta da un gruppo di cinque complessi enzimatici (Fig.1.1) presenti sulla membrana interna, rispettivamente: Complesso I o *Coenzima Q reduttasi*, Complesso II *Succinato Deidrogenasi*, Complesso III *Citocromo c Reduttasi*, Complesso IV *Citocromo c Ossidasi*, Complesso V *ATP sintasi*. I cofattori ridotti: NADH (Nicotinammide adenina dinucleotide) e FADH₂ (Flavina adenina dinucleotide) generati dal metabolismo intermedio di carboidrati grassi e proteine, sempre all'interno dei mitocondri, donano elettroni al Complesso I e II. Il trasporto degli elettroni avviene secondo gradiente elettrochimico e trasportati tramite *ubichinone* e *citocromo c* attraverso il Complesso III e IV. Quest'ultimo, dona un elettrone all'ossigeno portando così alla formazione di una molecola d'acqua. Il trasferimento degli elettroni è accoppiato col trasferimento di ioni H⁺ da parte dei complessi I, II e IV dalla matrice allo spazio intermembrana, generando di conseguenza sia un gradiente di PH, sia un potenziale di membrana che viene utilizzato dal

Complesso V per sintetizzare ATP a partire da Adenosina Difosfato (ADP) e fosfato inorganico. I complessi enzimatici possono trasportare gli elettroni grazie alla presenza di subunità all'interno della loro struttura, i Complessi I-III possiedono infatti un gruppo prostetico Fe-S, mentre il Complesso IV e *citocromo c* un gruppo eme.

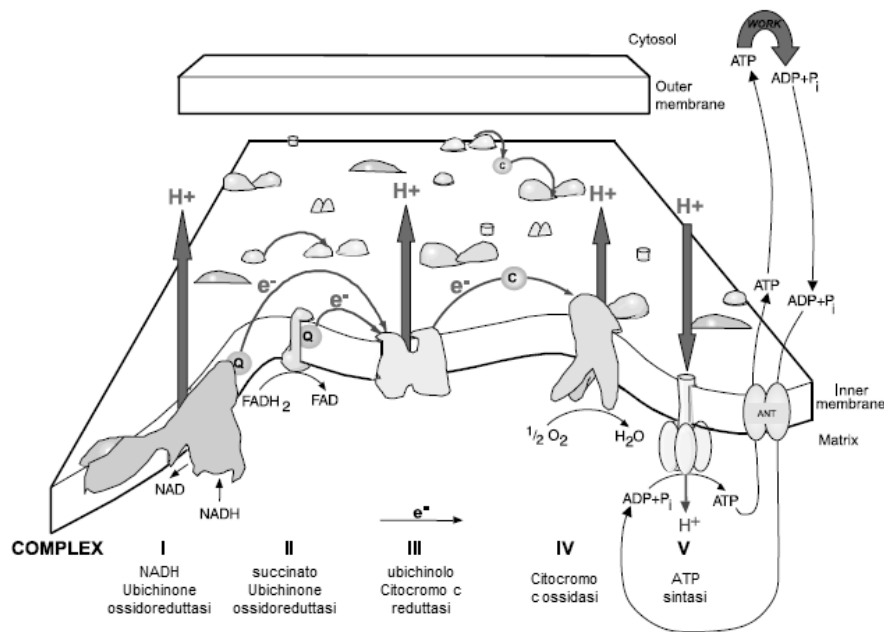


FIGURA 1.1. Gli elettroni donati dai cofattori NADH e FADH₂ prodotti dal metabolismo intermedio di carboidrati, proteine e grassi, passano attraverso i complessi I, II, III, IV per un gradiente elettrochimico trasportati da Ubichinone e Citocromo c. Il complesso IV dona un elettrone all'ossigeno, comportando la formazione di una molecola d'acqua. Il trasporto di un H⁺ nello spazio intermembrana determina un gradiente protonico, generando un potenziale di membrana utilizzato dal complesso V per sintetizzare ATP.

Poiché la fosforilazione ossidativa nel mitocondrio si basa sulla rapida diffusione di ioni e substrati sui siti di trasporto o di reazione nella membrana mitocondriale interna, il numero e la forma delle giunzioni delle creste potrebbe quindi regolare il livello di fosforilazione. Allo stesso modo, anche la forma ed il volume delle creste mitocondriali potrebbero alterare sia la diffusione della *citocromo c* tra la matrice e lo spazio intermembrana, sia la frazione di proteina legata alla membrana interna, regolando quindi la quantità di reazioni redox che coinvolgono l'emoproteina e la sua accessibilità verso la membrana esterna. Pertanto, esistono considerevoli prove che la membrana mitocondriale intera sia una struttura dinamica, capace di cambiare rapidamente forma in risposta alle alterazioni metaboliche o legate all'osmosi [15] diventando parte integrante dei meccanismi di feedback con

cui i mitocondri rispondono ai cambiamenti ambientali [16]. La fosforilazione ossidativa non è l'unico processo del metabolismo energetico della cellula in cui i mitocondri ricoprono un ruolo fondamentale. Vanno infatti ricordati anche il ciclo dell'urea, la beta-ossidazione degli acidi grassi e naturalmente, il loro coinvolgimento nell'omeostasi cellulare con la diretta partecipazione nel metabolismo di aminoacidi, lipidi, del colesterolo, steroidi e nucleotidi, così come nel ciclo di Krebs, nel ciclo degli acidi tricarbossilici e nella segnalazione cellulare ed apoptosi. E' stato osservato inoltre come i mitocondri abbiano anche un ruolo chiave nei primi stadi degli eventi che conducono all'invecchiamento [17], alcune prove indicano che questo meccanismo sia legato all'accumulo di mitocondri non funzionali che producono una significativa quantità di radicali ossigeno (ROS) [18]. E' stato quindi osservato come i mitocondri differiscano tra loro per morfologia, grandezza, forma delle creste [19], allo stesso modo anche il numero può variare per cellula: da poche centinaia a poche migliaia a seconda della tipologia cellulare, dalla sua specializzazione, così come tra diversi tessuti, organi ed organismi viventi [20] e si può avere in base ai diversi cambiamenti fisiologici oltre che alla variazione delle richieste di energia. Quando infatti la cellula avverte una diminuita funzionalità del mitocondrio a seguito, ad esempio di stress ambientali, alcuni segnali trasmessi al nucleo come la presenza di perossidi (H_2O_2) fanno sì che la cellula smetta di proliferare inducendo la proliferazione mitocondriale e l'amplificazione del mtDNA per produrre un maggiore numero di mitocondri funzionali così da ottenere maggiore energia per la riparazione del danno subito. A riparazione ultimata, la cellula può quindi rientrare nel ciclo cellulare e completare la crescita. Tuttavia, quando il danno persiste oppure risulta troppo grave per essere riparato, il mitocondrio può condurre la cellula verso il processo apoptotico. L'abbondanza di mitocondri per cellula quindi è strettamente controllata dall'attivazione di specifici pathway molecolari e fattori di trascrizione che si ipotizza vengano anche utilizzati in risposta allo stress ossidativo [21;22]

1.2 *Genoma mitocondriale*

Il genoma mitocondriale possiede alcune caratteristiche peculiari che lo rendono differente dal genoma nucleare. Una di queste consiste nell'impiego di un diverso codice genetico; ad esempio la tripletta UGA invece che fungere da codone di terminazione, come nel DNA nucleare, codifica per il triptofano. Le triplette AUA, AUC e AUU sono impiegati come codoni di inizio mentre AGA e AGG come codoni stop. Altre differenze riguardano la sua esclusiva trasmissione lungo la linea materna, il suo elevato tasso di mutazione, la poliploidia e la sua particolare organizzazione strutturale ed espressione.

Altamente conservata tra i mammiferi, la struttura molecolare consiste in un doppio filamento di forma circolare, della lunghezza di 16.569 bp il cui peso corrisponde a circa 10.000.000 Da [23]. Il numero totale di nucleoidi nel DNA mitocondriale può tuttavia cambiare, a causa di piccole mutazioni originate da inserzioni o delezioni. Un esempio, è la regione compresa tra le posizioni 513 e 524 del mitogenoma caratterizzata dalla presenza di ripetizioni dinucleotidiche AC che possono variare da un minimo di tre ad un massimo di sette ripetizioni [24]. Entrambi i filamenti di DNA presentano ulteriori caratteristiche quali: un'asimmetrica distribuzione dei nucleotidi che permette di distinguere una catena "leggera" (*light chain L*), ed una catena "pesante" (*heavy chain H*) caratterizzata da un maggiore contenuto di guanina che possiede il più alto peso molecolare. Inoltre, le catene presentano una conformazione superavvolta e risultano scarsamente associate a proteine [25]. Come nel DNA batterico, anche quello mitocondriale è organizzato in complessi nucleoproteici definiti nucleoidi, associati alla membrana mitocondriale interna oppure in forma libera [26]. All'interno dello stesso mitocondrio possono essere presenti una media di 4-5 copie del mitogenoma. Poiché ogni cellula può contenere centinaia di mitocondri, è stata stimata l'esistenza di migliaia di molecole di mtDNA totali [27; 28]

A livello funzionale, sulla regione codificante (*codR*) del DNA mitocondriale sono presenti 37 geni descritti nella Tabella 1.1, organizzati in forma compatta e senza introni [23], di cui: 2 geni

ribosomiali 12s e 16s (rRNA) , 13 geni essenziali per i polipeptidi della catena respiratoria che includono: sette geni, da ND1 a ND6 con ND4L codificanti per sette sub-unità del Complesso I; il gene *Cyt b* per il Complesso III; tre geni CO I, CO II e CO III codificanti per le sub-unità catalitiche del Complesso IV e due geni ATP6 e ATP8 per le sub-unità del Complesso V *ATP sintasi*. Infine, 22 geni RNA transfer (tRNA). La catena pesante H codifica per la maggior parte delle informazioni, su questo filamento sono infatti presenti i 2 geni rRNA, 14 tRNA e 12 su 13 geni per i polipeptidi, mentre i geni rimanenti sono localizzati sulla catena leggera L (Fig.1.2).

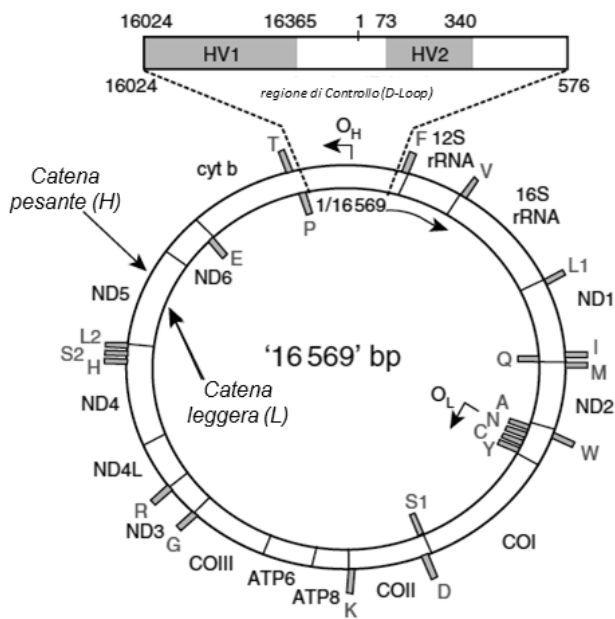


FIGURA 1.2. Rappresentazione schematica della struttura del genoma mitocondriale e dei relativi geni all'interno.

All'interno del genoma è anche presente una regione non codificante di 1100 bp che si estende dalla posizione 16024 alla posizione 576, definita anche regione di controllo (*control region* o CR), con funzione regolatoria della replicazione e della trascrizione tramite due specifici promotori al suo interno, uno per la catena pesante H (HSP) ed uno per la catena leggera L (LSP). Sebbene in letteratura, la regione di controllo venga definita anche

D-Loop (*Displacement Loop*), il termine utilizzato non è corretto. Quest'ultimo in realtà consiste in una struttura originata da terzo filamento DNA incorporato nella CR, lungo circa 650 nucleotidi che rappresenta quindi solo una parte di tutta la regione di controllo. Il confronto delle sequenze della CR tra mtDNA umano e murino, hanno evidenziato la presenza di tre "blocchi" altamente conservati (CBS1, CBS 2 e CBS 3) [29]. Sempre al suo interno, sono presenti inoltre tre piccole regioni le cui sequenze risultano altamente variabili a livello popolazionistico e per questo definite come: *Hyper Variable Region* (HVS), (Fig.1.3), organizzate dalla comunità forense rispettivamente in HVS I (16024 - 16365); HVS II (73 - 340); HVS III (438 - 576). La funzione delle regioni

ipervariabili risulta tutt'ora sconosciuta anche se potrebbero avere un ruolo importante nella trascrizione e replicazione del genoma mitocondriale. La sintesi del genoma inizia a partire

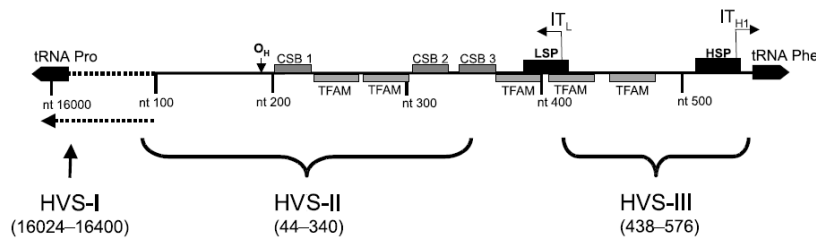


FIGURA 1.3. Rappresentazione schematica della regione di controllo e delle Hyper Variable Region I, II e III.

dall'origine di replicazione per la catena pesante H (O_H), a valle di LSP nella CR, procedendo lungo la catena leggera L per sintetizzare così la catena pesante H. Quando la sua replicazione raggiunge il

punto d'origine per la catena leggera L (O_L), presente all'interno di una seconda regione non codificante di soli 30 nucleotidi, la catena pesante H viene dislocata scoprendo il sito iniziatore per la sintesi della catena leggera L, che viene replicata in direzione opposta [30].

Posizione	Filamento	Abbreviazione	Descrizione	Taglia (bp)
16024-16569; 1-576		D-Loop	Regione di controllo	1122
16104-16569; 1-191		OH	Origine di replicazione	658
16158-16172			Segnale di terminazione	15
531-568			Promotore di trascr. H	38
577-647	H	F	tRNA Phenilalanina	71
648-1601	H	12S	12S rRNA	954
1602-1670	H	V	tRNA valina	69
1671-3229	H	16S	16s rRNA	1559
3230-3304	H	L1	tRNA leucina 1	75
3305-4263	H	ND1	NADH deidrogenasi 1	959
4263-4331	H	I	tRNA isoleucina	69
4329-4400	L	Q	tRNA glutammina	72
4401		-	Non codificante	1
4402-4469	H	M	tRNA metionina	68
4470-5511	H	ND2	NADH deidrogenasi 2	1042
5512-5579	H	W	tRNA triptofano	68
5580-5586		-	Non codificante	7
5587-5655	L	A	tRNA alanina	69
5656		-	Non codificante	1
5657-5729	L	N	tRNA asparagina	73
5730-576		O_L	origine di replicazione	31
5761-5826	L	C	tRNA cisteina	66
5826-5891	L	Y	tRNA tirosina	66

Posizione	Filamento	Abbreviazione	Descrizione	Taglia (bp)
5892-5900		-	Non codificante	9
5901-7445	H	COI	Citocromo ossidasi I	1545
7445-7516	L	S1	tRNA serina 1	72
7517		-	Non codificante	1
7518-7585	H	D	tRNA acido aspartico	68
7586-8294	H	COII	Citocromo c ossidasi II	709
8295-8364	H	K	tRNA lysina	70
8365-8572	H	ATP8	ATP sintasi 8	208
8527-9207	H	ATP6	ATP sintasi 6	681
9207-9990	H	COIII	Citocromo c ossidasi	784
9991-10058	H	G	tRNA glicina	68
10059-10404	H	ND3	NADH deidrogenasi 3	346
10405-10469	H	R	tRNA arginina	65
10470-10766	H	ND4L	NADH deidrogenasi 4L	297
10760-12137	H	ND4	NADH deidrogenasi 4	1378
12138-12206	H	H	tRNA istidina	69
12207-12265	H	S2	tRNA serina 2	59
12266-12336	H	L2	tRNA leucina 2	71
12337-14148	H	ND5	NADH deidrogenasi 5	1812
14149-14673	L	ND6	NADH deidrogenasi 6	525
14674-14742	L	E	tRNA a. glutammico	69
14743-14746		-	Non codificante	4
14747-15887	H	Cyt b	Citocromo b	1141
15888-15953	H	T	tRNA treonina	66
15954		-	Non codificante	1
15955-16023	L	P	tRNA prolina	69

TABELLA 1.1. Geni mitocondriali, posizioni e relative descrizioni.

1.3 Ereditarietà

E' ampiamente accettato che in tutti i mammiferi il DNA mitocondriale venga ereditato secondo la genetica non Mendeliana, esclusivamente attraverso il modello uniparentale della linea germinale materna [31;32]. Poichè è necessario che solo uno dei due gameti trasmetta il proprio pool mitocondriale allo zigote, è stato ipotizzato che questo meccanismo sia reso possibile tramite dimorfismo dei gameti maschili e femminili (anisogamia), caratterizzato da una ridotta dimensione degli spermatozoi ed una maggiore dimensione degli ovociti [33]. Durante la fecondazione, i mitocondri paterni coesistono inizialmente con quelli materni. Ciononostante, vengono eliminati in seguito e mai trasmessi nella generazione successiva [34]. Per spiegare questo meccanismo sono state create due teorie a riguardo: la prima ipotizza una “*diluizione semplice*”, dove l’mtDNA paterno presente in minor numero di copie viene semplicemente diluito dall’eccesso dei mitocondri materni all’interno dell’ovocita, e la sua presenza nella prole risulta difficilmente rilevabile [35], la seconda invece, riguarda la teoria di un meccanismo a “*degradazione attiva*” in cui il DNA mitocondriale paterno oppure i mitocondri, vengono selettivamente degradati prima o dopo la fecondazione per prevenire attivamente la trasmissione alla prossima generazione. Recentemente, diverse prove sono state osservate durante i primi stadi dell’embriogenesi a favore di questo meccanismo, nella maggior parte dei mammiferi inclusi gli esseri umani [36;37]. In particolare, Sutovsky e collaboratori [38] nei modelli di scimmia Rhesus, bovino e topo, hanno infatti osservato come all’interno del citosol dell’ovocita fecondato, i mitocondri paterni associati ad ubiquitina, un marcatore universale per il riciclaggio delle proteine, vadano in contro a proteolisi. La stessa proteina è stata inoltre rilevata a livello dell’apparato riproduttivo maschile, suggerendo l’ipotesi che i mitocondri paterni siano già marcati, così da comportare la loro selettiva degradazione ancora prima della fecondazione. Inoltre, la presenza di proibitina, una proteina situata nella membrana mitocondriale interna, è stata identificata come potenziale bersaglio per l’ubiquitinazione dei mitocondri negli spermatozoi [39], per tali motivi è ragionevole pensare che l’ubiquitinazione dei

mitocondri paterni sia responsabile della loro selettiva degradazione, sebbene questa teoria debba ancora essere verificata a livello funzionale. In generale risultano ancora poco chiare le ragioni che possano spiegare la presenza dei meccanismi atti ad impedire la trasmissione del mtDNA paterno, quella maggiormente accreditata riguarda la prevenzione di una possibile competizione tra i diversi DNA mitocondriali che potrebbe altrimenti condurre ad un conflitto tra i due genomi, con conseguenze letali [40]. Una seconda teoria alternativa vede l'adattamento all'anisogamia, come un sistema per escludere il mitogenoma paterno danneggiato a causa dell'intensa attività respiratoria nello spermatozoo[41].

L'analisi della trasmissione del mitogenoma nei *pedegree* animali ha evidenziato dei cambiamenti nel genotipo mitocondriale nelle diverse generazioni. Questo fenomeno è stato documentato in maniera particolare da Koehler *et al.*[42] nei bovini di razza Holstein, in cui la prole di una madre che possedeva contemporaneamente due aplotipi mitocondriali distinti in una condizione definita eteroplasmica, presentava entrambe le varianti in proporzioni diverse sia tra i componenti, sia tra prole e madre. Queste osservazioni portarono alla teoria, ampiamente accettata anche per gli esseri umani, dell'esistenza di un effetto collo di bottiglia (*bottleneck*) durante l'ovogenesi, attraverso cui solo una piccola frazione di molecole di mtDNA viene rappresentata nell'ovocita maturo [43; 44]. Durante la maturazione dell'ovocita si ha un consistente aumento del numero dei mitocondri e delle molecole di mtDNA, che raggiungono una quantità superiore alle centomila copie nella metafase II dell'ovocita, in preparazione alla richiesta metabolica della cellula al momento della fecondazione. Si presume che il collo di bottiglia possa verificarsi prima di questa fase di rapido accumulo di mtDNA attraverso due meccanismi definiti: *bottleneck* numerico, originato da una rapida segregazione degli aplotipi negli ovociti nei primi stadi dell'ovogenesi, attraverso la riduzione del numero di molecole di DNA mitocondriale per ciascuna cellula e da una deriva genetica stocastica. Diversi studi quantitativi per l'mtDNA hanno tuttavia presentato risultati sia a favore [45], sia contrari [46] allo sviluppo di una drastica contrazione nel numero di molecole per questo modello.

Un'ipotesi alternativa riguarda invece la replicazione preferenziale di una sottopopolazione del mitogenoma piuttosto che un'altra [47], risultando quindi solo in un apparente collo di bottiglia.

Due filoni di ricerca condotta sui modelli murini [48; 49; 50] hanno inoltre evidenziato la presenza di una selezione purificatrice per eventuali mutazioni patogene, che avrebbe luogo dopo il collo di bottiglia e la deriva genetica nella linea delle cellule germinali primordiali femminili, aiutando così a mantenere la funzionalità del DNA mitocondriale negli ovociti e di conseguenza anche nella prole. Tuttavia, i meccanismi molecolari alla base di questo evento, sono ancora poco conosciuti. Esperimenti condotti su scimmie Rhesus hanno suggerito l'esistenza di un ulteriore secondo *bottleneck* nei primi stadi della embriogenesi, prima della ripresa della replicazione del genoma mitocondriale [51]. È stato infatti osservato che quando ovociti eteroplasmici venivano fecondati, i singoli blastomeri mostravano un'ampia gamma d'eteroplasmia. Tuttavia, col procedere dello sviluppo embrionale, il feto derivante mostrava al contrario un ritorno ad una condizione omoplasmica, suggerendo quindi che nel feto le cellule somatiche possono andare incontro ad un collo di bottiglia del mtDNA anche dopo la formazione della blastocisti, mentre le cellule germinative femminili conservano comunque un certo livello di eteroplasmia, portando alla conclusione che le cellule somatiche e cellule germinali sono differenzialmente regolate durante l'embriogenesi [52].

1.4 Eteroplasmie

L'eteroplasmia viene definita come la presenza di due o più varianti di mtDNA all'interno della stessa cellula o mitocondrio [53]. In ambito medico, la loro origine è stata ampiamente investigata per fare luce sulle patologie mitocondriali, mentre in ambito popolazionistico e forense ricoprono un ruolo molto importante introducendo variabilità e per essere importanti marcatori nei casi di parentela.

Sebbene inizialmente si credesse che l'eteroplasmia fosse associata a stati patologici [54], recenti studi hanno dimostrato il contrario e con l'introduzione del *Next Generation Sequencing* la presenza delle eteroplasmie in individui sani è stata definitivamente confermata [55].

Data l'assenza di ricombinazione, l'unico meccanismo in grado di creare un'abbondante variabilità interindividuale del mitogenoma, è la presenza naturali processi di mutazione causati da sostituzione, inserzione e delezione di nucleotidi. Le cause presenti alla base dell'eteroplasmia possono essere trovate in primo luogo nel fatto che: il DNA mitocondriale contrariamente al DNA nucleare non è avvolto alle proteine istoniche che gli forniscono protezione, inoltre, la minore probabilità di riparare gli errori da *mismatch* durante la replicazione e la sua localizzazione vicino ai siti della fosforilazione ossidativa sulla membrana interna, lo espongono ad alti livelli di mutabilità per deaminazione sulla catena pesante, causata dalla presenza dei radicali ossigeno. Per queste ragioni il tasso di mutazione stimato nell'uomo corrisponde approssimativamente al valore di 6×10^{-7} , venticinque volte più rapido rispetto al DNA nucleare [56; 57]. Inoltre, la regione di controllo possiede un tasso di mutazione 10 volte maggiore rispetto alla regione codificante del mitogenoma, questo perchè le mutazioni presenti sulla CR, che non codifica per alcun gene, possono accumularsi senza essere soggette a selezione.

La quasi totale presenza della condizione omoplasmica nell'individuo, è determinata dalle forze di selezione e *bottleneck* durante l'ovogenesi, che agiscono sulla popolazione totale delle molecole di mtDNA presenti nella cellula, di conseguenza, la maggior parte delle eteroplasmie vengono perse nel corso di poche generazioni, comportando quindi la perdita di molte mutazioni *de novo*. Tuttavia, l'eteroplasmia può essere osservata sia nelle cellule germinali: in cui la madre trasmette diverse popolazioni di genoma mitocondriale alla generazione successiva; sia nelle cellule somatiche dove la replicazione del mitogenoma e la mitosi, originano diverse popolazioni cellulari caratterizzate da livelli variabili di eteroplasmia come osservato in diversi tessuti quali: muscolo, fegato, cervello, cuore, capelli e sangue [58]; dove il rapporto tra variante maggiore e minore è inversamente proporzionale all'aumentare del *turnover* cellulare [59].

Solamente una piccola percentuale di mutazioni lungo le cellule germinali riescono a fissarsi all'interno delle genealogie passando quindi da uno stato eteroplasmico ad una apparente condizione omoplasmica, come riportato in letteratura [60;61]. E' possibile supporre quindi, che la selezione giochi un ruolo primario sulla comparsa di nuove mutazioni prima che queste si propaghino fino ad essere individuate come eteroplasmie nel tessuto o nell'individuo, ciò può essere dovuto al fatto che alcune mutazioni non siano sufficientemente benigne per fissarsi stabilmente all'interno del mitogenoma. In alternativa, il *pattern* delle mutazioni si basa piuttosto su una differente propensione nella frequenza degli errori di replicazione a seconda del nucleotide presente[62]. Sono presenti due tipi di eteroplasmia lungo il genoma mitocondriale: si intende eteroplasmia di lunghezza (LHP) la variante caratterizzata dalla presenza di inserzioni nucleotidiche in prossimità dei tratti omopolimerici di citosina definiti anche *C-stretches*, descritte per la maggior parte all'interno della regione di controllo nelle sequenze ipervariabili HVS I, HVS II e III [63; 64;

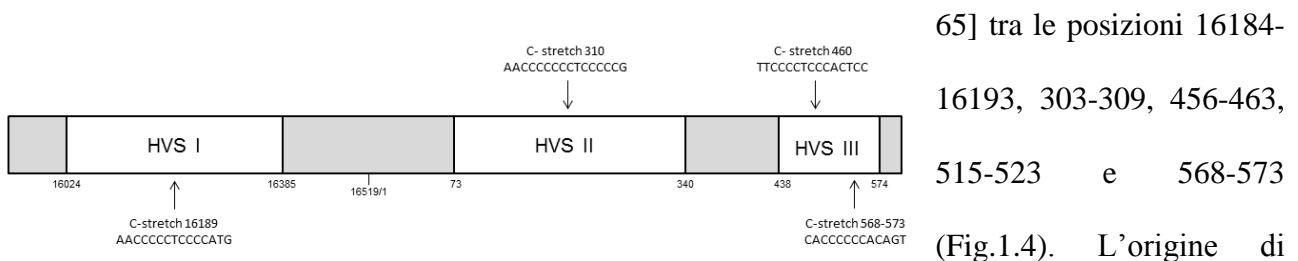


FIGURA 1.4. Le tre regioni ipervariabili all'interno della D-Loop e le rispettive sequenze di omopolimeri di citosina (*C-stretches*).

65] tra le posizioni 16184-16193, 303-309, 456-463, 515-523 e 568-573 (Fig.1.4). L'origine di queste eteroplasmie è in

stretta relazione tra la struttura delle *C-stretches* all'interno delle sequenze ipervariabili, e l'introduzione delle eteroplasmie di lunghezza. Forster *et al.* [64] hanno infatti osservato una maggiore probabilità nel verificarsi dell'evento quando la sequenza raggiunge il numero critico di otto nucleotidi ripetuti. Queste varianti possono essere responsabili inoltre di eventi di *slippage* dei filamenti di DNA durante la reazione di PCR [66], che causano l'insorgenza di artefatti analitici, aggiungendo una maggiore difficoltà interpretativa del dato ottenuto dal sequenziamento. Di conseguenza, dal momento che l'analisi delle LHP è strettamente dipendente dalla tecnologia impiegata per il sequenziamento, i dati ottenuti potrebbero non necessariamente rappresentare le varianti di lunghezza e la loro proporzione all'interno di un tessuto. Come risultato di questa

diversità, le inserzioni e le delezioni presenti all'interno dei tratti di policitosina vengono ignorati durante l'analisi forense nel confronto diretto dei profili genetici o tramite l'utilizzo dei database.

L'eteroplasmia puntiforme (PHP) consiste invece nella presenza di due nucleotidi nello stesso sito, esiste anche una condizione più rara in cui è possibile osservare eteroplasmia in due siti nello stesso individuo, e definita col termine di triplasmia [55]. Studi passati hanno evidenziato l'esistenza di alcune posizioni lungo il mitogenoma caratterizzate da un tasso di eteroplasmia maggiore rispetto ad altre [62; 67; 68; 69], e definite per tale ragione *hotspot*: come descritto per le posizioni 16093, 16029, 16153, 16189, 16192, 16293, 16309 e 16337 nella HVS I, e le posizioni 73, 152, 189, 207 e 279 nella HVS II [55;70; 71]. Per alcune di queste posizioni si osserva un elevato tasso evolutivo, ma tutti gli *hotspot* possiedono comunque un tasso di mutazione approssimativamente undici volte più veloce della media [62]. Forster *et al.* [72] riportano un aumento del tasso di di mutazione, e quindi una maggiore frequenza di eventi eteroplasmici, anche negli individui esposti alla radioattività naturale in cui siti maggiormente soggetti a PHP mostravano infatti un tasso di mutazione circa dodici volte più rapido rispetto a quello osservato per il DNA mitocondriale di derivazione Europea. Allo stesso modo, si osservano eteroplasmie correlate all'età ed invecchiamento, in base al gruppo popolazionistico od il genere. Tuttavia queste osservazioni descritte in letteratura risultano ancora controverse in quanto sono ancora necessari ulteriori approfondimenti [61; 73; 74]. Le PHP rappresentano uno strumento validissimo specialmente nel campo forense, nei casi in cui sia necessario ricostruire la parentela di individui presumibilmente appartenenti alla stessa linea materna, poiché la stessa posizione eteroplasmica osservata in due individui aventi lo stesso aplotipo e presente anche in diversi rapporti tra variante maggiore e minore, rappresenta una forte prova a favore dell'associazione di parentela.

Capitolo 2: Applicazioni del DNA mitocondriale

2.1 *Il DNA mitocondriale nello studio di popolazione e la sua filogenesi*

L'assenza di ricombinazione all'interno del mitogenoma e l'esclusiva ereditarietà lungo la linea materna, permettono un naturale raggruppamento di sequenze, formate da una serie di più mutazioni definite aplotipi, all'interno di principali cladi monofiletici conosciuti anche col termine di aplogruppi. Ciascun aplogruppo possiede al suo interno una sequenza specifica proveniente da una condivisione con un antenore comune, studi di filogeografia hanno dimostrato che gli aplogruppi del DNA mitocondriale possono essere associati ad ampie regioni geografiche, ed in particolare anche a specifici gruppi etnici [75;76]. Nei primi anni del 1980, la maggior parte delle analisi a livello popolazionistico condotte sull'intero mitogenoma consisteva nell'impiego di non più di sei enzimi di restrizione (RFLP) definiti *low resolution* [77], la distribuzione degli aplotipi risultava in questo caso rappresentata da una singola linea centrale presente a livello globale ritenuta essere l'aplotipo ancestrale, da cui si distaccano ramificazioni spesso regione-specifiche. Inoltre, la sua particolare distribuzione viene interpretata in supporto alla teoria del modello multiregionale alla base dell'origine dell'uomo [78]. A metà anni ottanta si passa all'impiego di un pool più esteso, *high resolution*, che comprende dai dodici ai quattordici RFLP [79]. Studi iniziali basati sull'uso questi enzimi verso numero ristretto di siti polimorfici, hanno indicato inizialmente il continente Asiatico come punto di origine del DNA mitocondriale [80]. Successivamente un più ampio studio condotto su 195 siti polimorfici di RFLP lungo l'intera sequenza del mitogenoma, analizzando 145 placente umane e due linee cellulari provenienti da cinque distinti gruppi popolazionistici, ha suggerito che tutte le popolazioni considerate possono essere ricondotte ad un singolo antenore

femminile, meglio definita come “Eva mitocondriale”, localizzata nel continente Africano e risalente a 200mila anni fa [79]. Tuttavia le prime filogenie che riconducevano all’Africa non erano sufficientemente robuste per dominare su altre ipotesi alternative come appunto quella basata sul modello multiregionale. Con l’introduzione della PCR nel 1990 comincia un estensivo sequenziamento della regione ipervariabile HVS I e II sulla CR consentendo uno studio approfondito grazie ad un maggiore numero di individui a disposizione, provenienti da continenti diversi, col fine di determinare le origini umane in ogni maggiore area geografica e permettendo così una migliore organizzazione degli aplogruppi all’interno dell’albero filogenetico. Tuttavia le diverse ramificazioni che ne caratterizzano la struttura, possono risultare confuse data la presenza di mutazioni ricorrenti a causa dei siti a rapida evoluzione presenti all’interno delle HVS I e HVS II. Il primo studio condotto su larga scala a livello popolazionistico da cui ha origine la prima organizzazione degli aplogruppi e la nomenclatura oggi universalmente accettata, ha interessato i nativi Americani in relazione agli eventi migratori ancestrali provenienti dal continente Asiatico, individuando quattro rami primari definiti in ordine alfabetico A, B, C, D [75]. Allo stesso modo altre linee filogenetiche con gli anni a seguire vengono caratterizzate per la popolazione Caucasica e Asiatica [81; 82], giungendo infine alla formulazione di regole cladistiche per l’ordine gerarchico degli aplogruppi e dei sub-aplogruppi stabiliti. Grazie alla successiva analisi dell’intero genoma mitocondriale che ha permesso un’elevata risoluzione molecolare, si è stabilito che le radici più profonde della filogenia del DNA mitocondriale sono riconducibili alla popolazione Africana [83] per la quale si possono definire sette linee principali da L0 a L5, localizzate nell’Africa sub-Sahariana caratterizzata infatti dalla più alta variabilità genetica e dalla presenza dei più antichi aplogruppi. Aplogruppi Africani al di fuori del continente sono estremamente rari e ristretti solo in particolari aree geografiche che subirono in passato flusso genico dall’Africa, come avvenuto per l’Europa mediterranea, Asia orientale e le Americhe. L’analisi ad alta risoluzione molecolare ha evidenziato infatti che i due terzi delle linee L presenti in Europa, sono state importate durante l’impero Romano, l’espansione Araba ed il commercio degli schiavi, mentre

solamente un terzo è stato introdotto durante le epoche preistoriche [84]. Circa ottanta-novanta mila anni fa a seguito di eventi di crescita e di colonizzazione della popolazione, in prossimità del corno d'Africa, l'aplogruppo L3 si scinde originando due sotto-aplogruppi: i macro-aplogruppi Euroasiatici M ed N, come evidenziato dalla loro significativa presenza in Etiopia [85]. Questo evento definito *Out of Africa* rappresenta l'origine di tutti aplogruppi non-Africani, risultando anche in una drastica riduzione della variabilità del genoma mitocondriale al di fuori del continente Africano. Successivamente, dall'aplogruppo N diverge una terza linea definita aplogruppo R. Tutti e tre gli aplogruppi costituiscono così la base di partenza per le successive linee diffuse in Eurasia, America ed Oceania quali: U, HV, JT, N1, N2 ed X presenti oggi in Europa, Sud-Est Asiatico e Nord Africa [86]; R5-R8, M2-M6 e m4'67 presenti in Asia meridionale [87]; gli aplogruppi A-G, Z ed M7-M9 sono invece diffusi in Asia orientale (Fig.1.5). Sebbene non sia facile associare l'origine di queste linee a specifici eventi preistorici o periodi, è possibile ipotizzare che la maggioranza degli aplogruppi comunemente presente in Europa, derivino tutti da eventi di ri-colonizzazione avvenuti dopo l'ultima glaciazione[86].

I più recenti studi filogeografici, genetica di popolazione e di antropologia molecolare si basano sull'utilizzo del sequenziamento di Sanger per studiare approfonditamente la struttura genetica delle popolazioni, la loro storia demografica e per ottenere maggiori dettagli sulla genesi degli eventi migratori ed i percorsi, che hanno avuto origine sia in epoca preistorica sia in corrispondenza di importanti eventi storici. In questi ultimi anni anche l'impiego delle *Next Generation Sequencing* nell'analisi dell'intero genoma mitocondriale ricopre un ruolo importante in questi settori, fornendo ad esempio informazioni riguardo i *pattern* migratori specifici legati al sesso come descritto in un recente studio condotto sulle popolazioni dell'arcipelago Indonesiano, che ha evidenziato una maggiore espansione geografica causata dall'effetto fondatore per il DNA mitocondriale, rispetto al cromosoma Y, suggerendo come in epoca preistorica, le donne fossero maggiormente propense rispetto agli uomini a migrare [88]; oppure fornendo informazioni sulla variabilità ed affinità genetica tra le popolazioni [89; 90]. Di grande rilevanza inoltre, è l'applicazione di questa nuova

metodica sul DNA mitocondriale antico estratto da reperti archeologici. A riguardo, l'analisi svolta su reperti ossei di epoca pre-Colombiana ha permesso di ricostruire la rotta migratoria per le Americhe e la sua collocazione temporale, rilevando inoltre la presenza di ulteriori aplogruppi e quindi una maggiore variabilità genetica, perduta in seguito ai colli di bottiglia introdotti dalle passate colonizzazioni Europee [91]. Sulla base di quanto riportato, l'analisi del mtDNA risulta ancora estremamente importante poiché la sua ereditarietà lungo la linea materna è tutt'ora uno strumento informativo per molti aspetti legati alla genetica delle popolazioni.

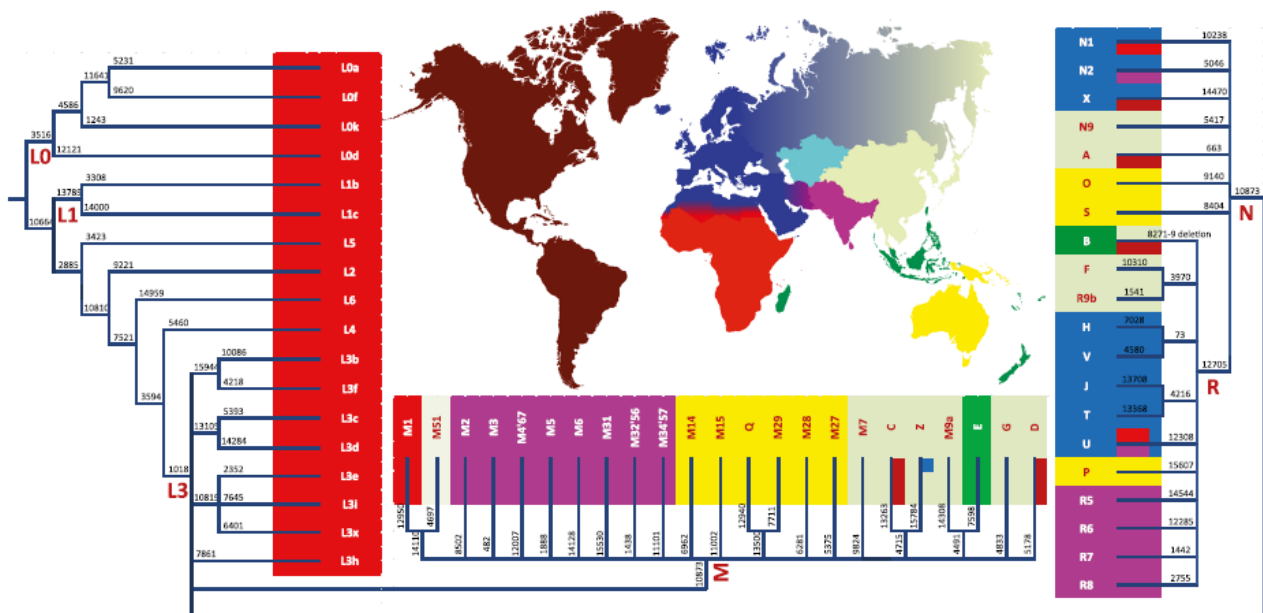


FIGURA 2.1. Distribuzione degli aplogruppi per il DNA mitocondriale. I colori rappresentano le principali caratteristiche di distribuzione geografica per ciascuna linea filogenetica.

2.2 *Il DNA mitocondriale in Genetica Forense, analisi ed applicazioni*

L'analisi del mitogenoma si dimostra un importantissimo strumento anche in genetica forense per calcolare le probabilità di una compatibilità tra due campioni biologici ai fini dell'identificazione personale o di una parentela. All'inizio degli anni novanta i genetisti forensi compresero l'importanza di analizzare il mitogenoma per sfruttare matrici biologiche caratterizzate da una bassa concentrazione di DNA nucleare, grazie soprattutto all'introduzione della *polymerase chain reaction* (PCR). I progressi condotti dalla filogenia e dalla filogeografia sulle varianti di mtDNA nelle popolazioni umane risultano importanti anche in questo settore per la condivisione delle stesse metodologie ed approcci analitici, come ad esempio quelli impiegati nell'analisi del DNA degradato sviluppati dagli studi di antropologia molecolare su DNA antico [92], così come la caratterizzazione delle linee mitocondriali per l'ancestralità delle popolazioni, che permette di determinare in questo campo, la provenienza etnica del donatore della traccia biologica, necessaria ai fini identificativi.

L'analisi genetico-forense del mitogenoma consiste nel confronto tra la sequenza del campione in esame, ottenuta sequenziando la CR attraverso il metodo di Sanger che rappresenta tutt'ora il *gold standard*, ed una sequenza nota utilizzata come riferimento definita anche sequenza di Anderson o *Cambridge Reference Sequence* (rCRS). Una delle maggiori problematiche che si incontrano a livello forense, è data dalle sequenze di omopolimeri di citosina definite *C-stretches* all'interno delle regioni ipervariabili, che oltre ad essere responsabili delle eteroplasmie di lunghezza possono introdurre nella fase di PCR che precede il sequenziamento, artefatti a danno dell'interpretabilità della sequenza. Poiché proprio da questa dipende il peso dell'evidenza in sede giudiziaria, la chiarezza del dato analitico risulta quindi uno degli obiettivi più complessi nel campo genetico-forense.

Una volta ottenute le sequenze del genoma analizzando entrambi i filamenti della molecola per ottenere il massimo *coverage* e quindi un migliore risultato analitico, il confronto dei profili

genetici dal DNA mitocondriale, svolto per stabilire a quali campioni appartenga la traccia biologica rinvenuta o la linea materna, può risultare in una esclusione: quando tra i campioni esaminati sono presenti due o più differenze nella sequenza; inconclusività: quando tra i campioni è presente una sola differenza; impossibilità di esclusione: in cui i campioni analizzati presentano la stessa sequenza o la stessa eteroplasmia di lunghezza. Nella valutazione di un profilo genetico condotto sul DNA mitocondriale è necessario tuttavia fare alcune considerazioni di carattere biologico riguardo l'eterogeneità del tasso di mutazione: è noto infatti che il livello della stabilità molecolare può variare lungo la molecola del genoma mitocondriale, come osservato per gli *hotspot* mutazionali riportati in letteratura [62], per cui una singola differenza tra due campioni dovrebbe essere valutata in base al tasso di mutazione specifico per la posizione considerata [93]. Riguardo le eteroplasmie puntiformi e di lunghezza, la loro presenza non può essere usata per stabilire un'esclusione nel confronto tra due tracce analizzate. Questo perché il genoma mitocondriale è caratterizzato sia da un elevato tasso di mutazione, sia da eventi di *bottleneck* durante l'ovogenesi che eliminano la maggior parte delle mutazioni *de novo*. La condivisione di un'eteroplasmia puntiforme lungo la linea materna, fornisce invece un ulteriore supporto per una non esclusione.

Nel caso di una non esclusione è necessario quindi stabilire la forza di associazione tra l'aplotipo osservato e quello appartenente alla traccia biologica attraverso una stima statistica per indicare la rarità dell'aplotipo tra individui non imparentati. A tal proposito, si applica il metodo della conta osservando quante volte l'aplotipo è presente all'interno di un database [94]. Poiché il DNA mitocondriale mostra alti livelli di variabilità, un database che include solamente cento individui non è sufficientemente rappresentativo dell'intera variabilità genetica nella popolazione, fatta eccezione per particolari gruppi che hanno subito un recente effetto fondatore, di conseguenza si possono verificare *bias* nella stima della frequenza dell'aplotipo per il campione analizzato, per cui, maggiore è il numero di individui non imparentati tra loro all'interno del database e più forte sarà il potere statistico. Solitamente i laboratori di genetica forense tendono ad evitare l'uso di database di libera consultazione presenti in letteratura, poiché spesso non rispondono agli standard di qualità,

prediligendo piuttosto quelli propri pubblicati sulle riviste forensi specializzate. Tuttavia come evidenziato da Forster et al. [95], nemmeno questi sono esenti da possibili errori. Un alternativa è quindi offerta dal database online EMPOP (<http://www.empop.org/>), sviluppato dall'Istituto di Medicina Legale e dall'Istituto di Matematica dell'Università di Innsbruck, rappresenta lo strumento più attendibile assicurando un alto standard di qualità grazie al suo costante aggiornamento. Inoltre EMPOP, a partire dall'aplotipo permette di inferire l'aplogruppo grazie l'algoritmo EMMA basato sul confronto di 14.990 genomi completi di DNA mitocondriale presenti nel database di sequenze genetiche GenBank[®] e di 3925 aplotipi virtuali ottenuti da Phylotree per rappresentare la base per l'elaborazione della stima dell'aplogruppo, assieme a 19.171 sequenze ottenute dalla regione di controllo utilizzate per determinare le fluttuazioni dei tassi di mutazione con il massimo approccio probabilistico [96].

A partire dagli anni duemila, la Società Internazionale di Genetica Forense (ISFG) in seguito alle problematiche precedentemente descritte, ha definito e pubblicato periodicamente linee guida specifiche per assicurare un elevato standard di qualità nell'analisi del genoma mitocondriale nei laboratori forensi e per l'interpretazione dei risultati. Nella versione più recente pubblicata nel 2014 [97], oltre alle raccomandazioni generali per una buona pratica in laboratorio che includono: appropriate condizioni lavorative, adozione di controlli positivi e negativi, sequenziamento di entrambe le catene del genoma, analisi di conferma indipendenti delle sequenze consenso e partecipazione regolare a programmi di test di competenza GEDNAP (*German DNA Profiling Group*) per l'analisi forense del DNA mitocondriale, viene sostenuta anche l'importanza del sequenziamento dell'intera regione di controllo rispetto alle sole regioni ipervariabili HVS I e II, per tutti gli studi di genetica di popolazione che sono stati sviluppati appositamente per allestire database a scopo forense. Vengono stabilite anche le regole per il corretto allineamento delle sequenze con la rCRS in accordo con la filogenia del DNA mitocondriale, così come l'esatta nomenclatura delle inserzioni e delezioni e quella secondo le convenzioni IUPAC (*International Union of Pure and Applied Chemistry*) per indicare i nucleotidi in eteroplasmia.

Uno dei punti importanti delle linee guida tutt'ora in fase di studio, riguarda l'aspetto legato all'interpretazione delle eteroplasmie puntiformi e di lunghezza, per le quali ciascun laboratorio deve tenere in considerazione i limiti imposti e la qualità del sequenziamento di Sanger. Dal punto di vista interpretativo, le linee guida suggeriscono per entrambe le eteroplasmie, che le differenze osservate non costituiscono una prova valida per escludere un confronto positivo tra due aplotipi. La stessa eteroplasmia puntiforme presente in due aplotipi identici, potrebbe al contrario accrescere la forza d'associazione tra due individui. Per quanto riguarda le eteroplasmie di lunghezza invece, poiché è nota la loro variabilità a livello dei tratti di policitosina (§1.4) definiti *C-stretches*, ai fini interpretativi queste regioni vengono ignorate nella comparazione forense. Al contrario, per i database genetici a scopo forense, le eteroplasmie di lunghezza vanno riportati nominando la variante dominante. In quest'ultima versione, viene inoltre fornito un corretto approccio statistico in base al database di popolazione utilizzato ed i criteri di ricerca da adottare all'interno dello stesso così da evitare la presenza di *bias* associando inoltre l'uso di EMPOP per una migliore stima della rarità dell'aplotipo.

Il caso più famoso dell'applicazione del DNA mitocondriale nella genetica forense ai fini identificativi, riguarda il riconoscimento nel 1994 della famiglia Reale russa, assassinati nel 1918 in Siberia, nella città di Ekaterinburg, da un plotone di esecuzione durante la rivoluzione d'ottobre. Benché i fatti storici e gli esami medico-legali svolti sui reperti ossei potessero far supporre che i resti trovati appartenessero alla famiglia reale Russa, solamente l'analisi genetica attraverso il DNA mitocondriale poteva fornire la piena certezza dell'identità dei corpi ritrovati. La comparazione delle sequenze delle regioni ipervariabili HVS I e II del genoma mitocondriale appartenente a due discendenti provenienti dalla stessa linea mitocondriale della nonna materna dello Zar, ed al presunto corpo di Nicola II, ha rivelato un match per tutte le posizioni analizzate ad eccezione della posizione 16169, che in quest'ultimo era presente in forma eteroplasmica in un rapporto tra citosina e timina di 3,4:1. Sebbene questo *mismatch* potesse essere interpretato come un dato controverso nel stabilire l'autenticità, un decisivo supporto venne invece fornito attraverso l'analisi del

mitogenoma appartenente al fratello dello Zar Georgij Romanov, morto nel 1899, che presentava all'interno del profilo la stessa posizione eteroplasmica C/T differente solo nel rapporto tra le due basi ed assente invece nel profilo di un terzo discendente vivente sulla stessa linea materna. Alla luce di questi dati, la probabilità di osservare un eteroplasmia nella stessa posizione in due individui non imparentati è pari a 2.4×10^{-4} , inoltre, il rapporto di verosimiglianza ottenuto tra lo Zar Nicola II ed il fratello risulta 150 volte maggiore rispetto all'ipotesi se i resti rinvenuti appartenessero ad una persona non imparentata. Di conseguenza non esiste alcuna ragionevole obiezione scientifica per rifiutare l'autenticità dei reperti all'ultimo sovrano della Russia.

Risulta quindi chiaro che questo marcatore di linea rappresenta nella genetica forense un validissimo strumento anche per ricostruire il legame familiare tra un individuo ed i discendenti, e per identificare vittime coinvolte in disastri di massa ed omicidi, dove i resti umani risultino degradati o rinvenuti solo molti anni dopo la morte e quindi impossibili da analizzare attraverso i convenzionali marcatori STRs.

2.3 *Il DNA mitocondriale in Genetica Medica e le patologie associate*

Grazie alle conoscenze acquisite in ambito genetico, è possibile studiare dal punto di vista biomedico, i meccanismi presenti alla base di patologie derivanti da mutazioni a carico del DNA mitocondriale. Oltre 300 polimorfismi a singolo nucleotide (SNPs) sono riportati ad oggi sul database MITOMAP (<http://mitomap.org>; ultimo accesso 22 Febbraio 2017) associati a mitopatie che possiedono caratteristiche uniche riguardo l'insorgenza, gravità ed ereditarietà. Poiché queste patologie sono spesso clinicamente eterogenee molti pazienti non rientrano all'interno di specifiche sindromi cliniche, tuttavia esistono in letteratura esempi noti di malattie mitocondriali quali: MIDD (*mitochondrially inherited diabetes and deafness*) [98], MELAS (*mitochondrial myopathy,*

encephalomyopathy, lactis acidosis, stroke-like symptoms) [99], MILS (*mathernally inherited Leigh's syndrome*) [100], MERRF (*myoclonic epilepsy with ragged red fibers*) [101] e LHON (*Leber's hereditary optic neuropathy*) [102]. Queste patologie sono responsabili di un ampio intervallo di fenotipi clinici che includono miopatie, difetti visivi e dell'udito, disordini neurologici, problemi respiratori, e sono per la maggior parte associate a mutazioni genetiche dell'apparato della fosforilazione ossidativa, come osservato per le mutazioni che coinvolgono i geni *MTND1*, *MTND2*, *MTND3*, *MTND4*, *MTND5* e *MTND6*, codificanti per le subunità proteiche del Complesso I, principalmente responsabili di LHON, MILS e MELAS [103], tuttavia le patologie mitocondriali possono insorgere anche da anomalie presenti sul DNA nucleare per i geni che codificano proteine coinvolte nel metabolismo del mitocondrio. Mutazioni a carico del genoma mitocondriale sono state inoltre trovate in associazione con diabete, autismo, cancro, Alzheimer e Parkinson [104; 105; 106; 107].

Data la natura poliploide del mtDNA le mutazioni patogene possono essere presenti in tutte le copie del genoma in condizione omoplasmica, oppure distribuite solo in alcune molecole in eteroplasmia come osservato per gravi mitocondriopatie infantili. Come descritto nei precedenti paragrafi, dal momento che la presenza di una variante del mtDNA all'interno del mitocondrio è influenzata da eventi di *bottleneck* genetico durante l'ovogenesi ed amplificata preferenzialmente per effetto fondatore, lo stesso può verificarsi nel caso di una mutazione patogena, che verrà trasmessa dalla madre alla generazione successiva. Poiché l'evento di *bottleneck* può determinare una rapida variazione della frequenza delle mutazioni patogene nel corso delle generazioni, risulta inoltre complicato prevedere l'espressione delle patologie mitocondriali nell'uomo. Generalmente le manifestazioni delle patologie mitocondriali insorgono negli individui solamente quando la variante patogena associata alla malattia supera una soglia pari al 70% del numero totale di molecole di mtDNA [108]. Tuttavia i diversi tessuti potrebbero essere più o meno propensi a manifestare la patologia [109], questo perché in primo luogo alcuni sono più dipendenti di altri alla fosforilazione ossidativa, come il tessuto nervoso, il tessuto cardiaco, il tessuto muscolare scheletrico e retina. In

secondo luogo, la frequenza iniziale della mutazione patogena varia da tessuto a tessuto a causa della segregazione del mtDNA *wild-type* e mutante durante l'embriogenesi. Il DNA mitocondriale, come descritto precedentemente, è caratterizzato da un'elevata variabilità grazie all'accumulo nel tempo di polimorfismi coinvolti anche nell'adattamento climatico e metabolico. Sebbene queste varianti abbiano permesso ai nostri antenati di adattarsi ai cambiamenti ambientali, è stato inoltre suggerito che queste potessero anche giocare un ruolo dannoso nelle patologie dell'uomo moderno legate alla bioenergetica [110; 111], e che siano quindi state trasmesse all'interno degli aplogruppi nelle popolazioni, come riportato da studi riguardanti suscettibilità del morbo di Parkinson o sul diabete di tipo 2 [112].

Le patologie mitocondriali, ad oggi non sono direttamente curabili nonostante siano stati sviluppati diversi approcci mirati a cambiare nei pazienti la quantità di eteroplasmia patogena, come ad esempio l'impiego di specifiche nucleasi progettate per "clivare" la molecola di mtDNA sul sito mutato. In assenza di terapie disponibili quindi, le strategie adottate con lo scopo di prevenire la trasmissione della patologia dalla madre alle generazioni successive, risultano estremamente importanti. A tal proposito esistono strategie classiche mirate a rimpiazzare l'ovocita dal paziente o monitorare l'eteroplasmia fetale nell'embrione. Le strategie moderne invece consistono nel rimpiazzare il genoma mitocondriale patogeno. Tuttavia in questo caso ci si domanda quale sia la magnitudo della segregazione dell'aplotipo di mtDNA non patogeno che può determinare una potenziale amplificazione della mutazione patogena oppure la sua definitiva rimozione [113].

Capitolo 3: Analisi del DNA mitocondriale

3.1 *Il Polimorfismo a singolo nucleotide*

Si definisce un polimorfismo a singolo nucleotide o SNP , la variazione di una singola base sul genoma presente in almeno 1% nella popolazione. Questi polimorfismi possono verificarsi sia in regioni non codificanti sia in quelle codificanti. In questo caso la mutazione viene definita sinonima, quando il polimorfismo non altera la struttura della proteina espressa e la propria funzionalità, non-sinonima, quando invece la mutazione altera la sequenza aminoacidica della proteina, risultando in un fenotipo alterato nell'uomo. A seconda che la sostituzione coinvolga una purina od una pirimidina si definisce transizione la sostituzione dello stesso genere (C/T o A/G), mentre si definisce trasversione il passaggio da una purina ad una pirimidina e viceversa. Appartengono alla categoria degli SNPs anche le inserzioni e le delezioni che coinvolgono singole basi. In generale, i polimorfismi a singolo nucleotide sono prevalentemente introdotti da errori di incorporazione nella sequenza del genoma durante la replicazione oppure da eventi di mutagenesi di composti chimici o radiazioni ionizzanti. Il tasso di mutazione a carico degli SNPs presenti sulla molecola del DNA mitocondriale, ed in particolare quelli presenti nella regione di controllo, risulta più rapido rispetto a quello osservato nel DNA nucleare, oltre ad essere caratterizzato da una eterogeneità nella frequenza a causa di *hotspot* a rapida evoluzione localizzati lungo tutto il mitogenoma. La frequenza media individuata per l'intera molecola è stimata pari al valore di 6×10^{-7} per la regione codificante, e 5×10^{-7} nelle regioni ipervariabili HVS I e II. L'elevato tasso di mutazione nel corso delle generazioni ha contribuito ad introdurre un'elevata variabilità a carico della sequenza del mitogenoma, e poiché questi polimorfismi si sono fissati nel genoma e trasmessi

sottoforma di aplotipo da generazione in generazione, sono quindi diventati i marcatori d'elezione sia nel campo forense, definiti in questo caso come marcatori biallelici, sia nell'ambito evoluzionistico per ricostruire la storia migratoria di una popolazione, l'ancestralità di un individuo e l'identificazione personale negli studi di parentela.

3.2 *Il sequenziamento di Sanger nell'analisi del DNA mitocondriale*

Il metodo basato sul sequenziamento di Sanger, sebbene sia stato descritto per la prima volta quasi trent'anni fa, rappresenta ancora oggi il *gold standard* nel forense e nella genetica medica diagnostica applicato all'analisi della sequenza del genoma mitocondriale. La tecnica sfrutta la reazione biochimica della replicazione del DNA e consiste in un primo passaggio di PCR mirato all'amplificazione della regione d'interesse. Gli ampliconi prodotti sono quindi utilizzati come stampo nella reazione di sequenziamento, che consiste in una seconda reazione di PCR in cui assieme ai nucleotidi per l'estensione della catena ed i primers specifici, sono aggiunti quattro di-deossi-nucleotidi trifosfato (ddNTPs) come terminatori di catena, ognuno marcato con uno specifico fluorocromo. La principale caratteristica dei ddNTPS è l'assenza del gruppo ossidrilico al carbonio 3', che impedisce l'incorporazione della successiva base durante la polimerizzazione della catena nucleotidica. Poiché la concentrazione dei dNTPs è più alta rispetto a quella dei di-deossi-nucleotidi, quest'ultimi vengono incorporati lungo la molecola ad intervalli casuali e questo produce un intervallo di frammenti sintetizzati di diverse lunghezze, che verranno analizzati con l'impiego di un sequenziatore automatico che applicando al campione un voltaggio elettrico, in un processo definito iniezione elettrocinetica che si origina attraverso un catodo (-) ed un anodo (+), si ottiene la separazione degli ampliconi in base alla loro dimensione specifica tramite un polimero viscoso di dimethyl-polyacrylamide in concentrazione al 4% oppure al 6% all'interno di un

capillare di vetro dalla dimensione di 36/80 cm e dal diametro di 50µm. Ad una maggiore lunghezza del capillare e concentrazione del polimero, corrisponde una maggiore separazione dei frammenti e di conseguenza una più alta risoluzione per quegli ampliconi che differiscono solo di poche paia-basi. Il DNA quindi si muove attraverso il polimero, dal catodo verso l'anodo, la corrente elettroforetica è garantita da un Buffer che fornisce ioni ai poli + e - la cui composizione può alterare la migrazione del DNA. Per fare in modo che gli ampliconi restino denaturati, al campione viene aggiunta un'aliquota di Formamide in soluzione, inoltre, per impedire che le molecole di DNA possano assumere conformazioni alterate dovute alle attrazioni intramolecolari durante il processo della corsa elettroforetica viene comunemente mantenuta una temperatura di 60°C. I frammenti che giungono in prossimità dell'anodo passeranno attraverso una sorgente luminosa fornita da un laser a ioni Argon (488nm), i frammenti più piccoli arriveranno quindi più rapidamente rispetto a quelli di dimensioni maggiori al punto di rilevazione, dove le differenti lunghezze d'onda emesse dai fluorofori corrispondenti a ciascun nucleotide, vengono infine registrate da un sensore a CCD camera e convertite in segnale analitico (Fig.3.1).

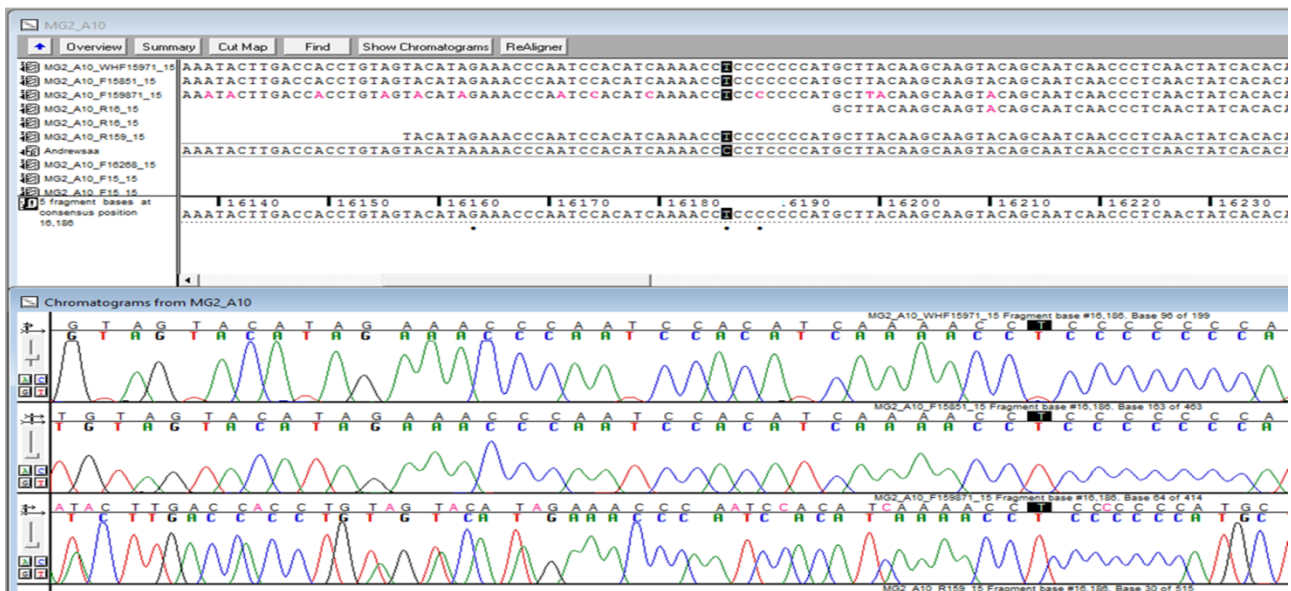


FIGURA 3. 1. Elettroferogramma della sequenza del genoma mitocondriale ottenuta tramite metodo Sanger

3.3 *Tecnologie di Next Generation Sequencing ed applicazioni*

Con il termine *Next Generation Sequencing* (NGS), si indica un insieme di tecnologie sviluppate per l'analisi di estese regioni genomiche attualmente impossibili o difficili da analizzare attraverso il metodo di Sanger, col principale vantaggio di ottenere un grande volume di dati a costi ridotti, sfruttando la combinazione di diverse strategie analitiche che includono: preparazione del template, sequenziamento, allineamento genomico. Nello specifico, il primo passaggio può avvenire tramite due distinti metodi: la prima consiste in una reazione di amplificazione clonale che può essere svolta in fase solida tramite primers legati covalentemente ad un substrato per ottenere clusters di ampliconi, oppure tramite l'allestimento di una *emulsion* PCR all'interno di un sistema separato, in cui ciascuna libreria genomica viene complessata ad una sequenza oligonucleotidica che funge sia da innesto per il *primer*, sia da adattatore per una biglia di amplificazione utilizzata da supporto ai filamenti di DNA neosintetizzati. Sebbene il metodo basato sull'amplificazione clonale offre il vantaggio di evitare la perdita di sequenze genomiche, alcuni protocolli risultano laboriosi e richiedono l'uso di una grande quantità di DNA (3-20µg). Il secondo metodo nella preparazione del template invece, oltre ad una minore quantità di DNA richiesto, non prevede l'allestimento di una reazione di amplificazione che può accidentalmente introdurre mutazioni nella sequenza amplificata a livello delle regioni ricche di G-C o A-T. Anche in questo caso il metodo utilizza una fase solida sulla quale sono fissati, a seconda della tecnologia scelta, adattatori, *primers* o molecole di DNA polimerasi, dove si lega il template precedentemente ridotto in frammenti di ~200 – 250 bp, in una reazione diretta di NGS.

Anche nella fase di sequenziamento sono stati sviluppati diversi metodi, un primo sistema viene chiamato: terminazione ciclica reversibile (CRT), e consiste in una serie di cicli in cui il terminatore di catena corrispondente alla base complementare del filamento stampo da sequenziare, una volta incorporato, subisce un clivaggio finale per rimuovere dalla struttura del nucleotide la porzione del

terminatore ed il marcatore fluorescente. Un secondo sistema definito sequenziamento per ligazione, prevede l'ibridazione ciclica sul frammento da analizzare con sonde complementari marcate con fluorocromo, che viene rimosso alla fine di ogni ciclo di ibridazione. Un terzo sistema associato all'amplificazione clonale di tutti i frammenti che compongono la libreria genomica, si basa sulla variazione elettrochimica tramite sensori semiconduttori presenti in serie all'interno di un chip invece che sull'utilizzo di un fluorocromo. La caratteristica di questo sistema, consiste nel fatto che l'incorporazione di un nucleotide, determina un aumento del PH a causa del rilascio di ioni H^+ da parte della DNA polimerasi, che viene registrato dai semiconduttori e convertito in segnale analitico corrispondente alla base annessa (Fig.3.2). La variazione del PH sarà quindi proporzionale al numero di basi complementari presenti sul filamento stampo da sequenziare. Il quarto ed ultimo passaggio delle metodiche di sequenziamento NGS è rappresentato dall'allineamento dei frammenti sequenziati (*reads*), che viene effettuato comunemente, attraverso l'uso di una sequenza di riferimento ottenuta da genomi noti. Un secondo metodo impiegato spesso nell'analisi del genoma batterico o nelle librerie BAC, consiste invece nell'assemblare le *reads de novo* per determinare l'esatta sequenza del genoma attraverso l'uso di un set di algoritmi [114].

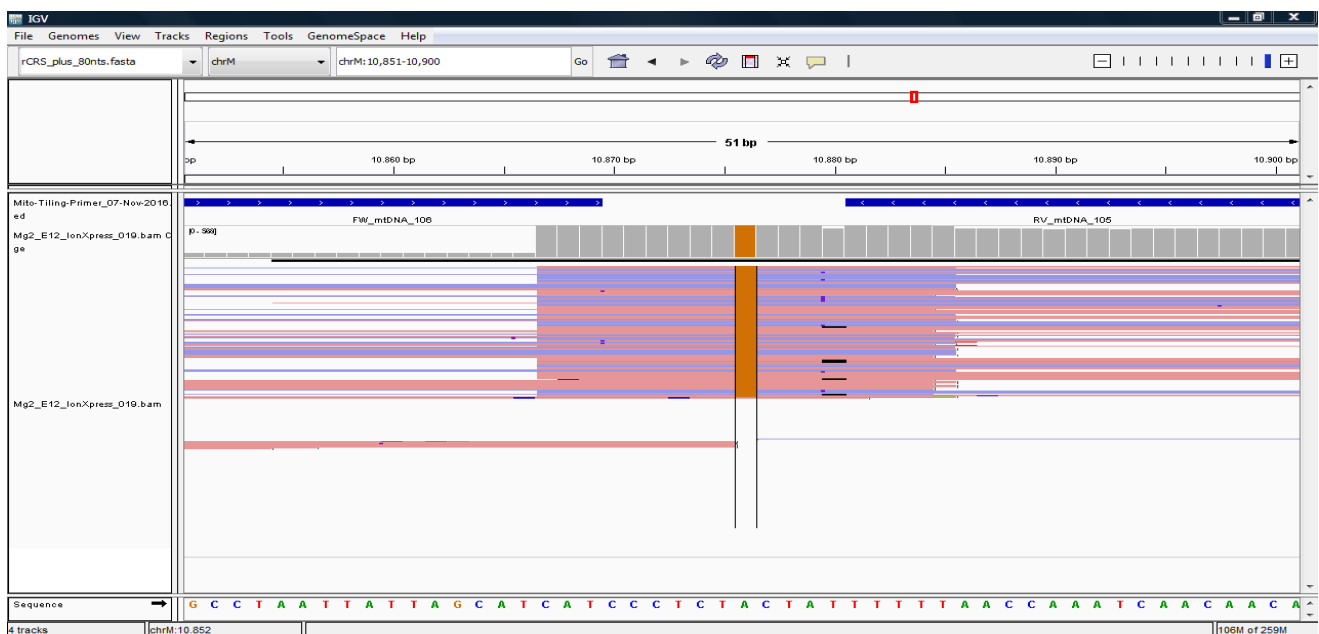


FIGURA 3. 2. Rappresentazione grafica delle differenze presenti nella sequenza del mitogenoma analizzato con la rCRS. Le posizioni variare sono evidenziate dal colore corrispondente alla base mutata

La grande mole di dati a basso costo prodotti dalle tecnologie *Next Generation Sequencing* hanno permesso in questi ultimi anni un'ampia versatilità nell'utilizzo di queste piattaforme nei più svariati ambiti di ricerca mirati allo studio dell'intero genoma umano, coinvolgendo la genetica medica nell'analisi delle sue varianti e la loro associazione ai fenotipi patogeni, a livello epigenetico, nella farmacogenetica nella scoperta di nuovi alleli, così come nella genetica di popolazione per i marcatori di linea del cromosoma Y e del DNA mitocondriale e per i geni autosomici. Infine un'ulteriore applicazione tutt'ora in fase di studio e validazione, riguarda l'uso del NGS nel settore della genetica forense, che attraverso l'analisi dell'intero genoma mitocondriale, troppo esteso per il solo metodo di Sanger, permetterebbe di fornire un aumento del potere discriminativo in individui che apparentemente risiedono sulla stessa linea materna.

3.4 *La sequenza di riferimento di Cambridge*

Nel 1981 viene sequenziata per la prima volta presso il laboratorio di Frederick Sanger a Cambridge l'intera sequenza del genoma mitocondriale da Anderson *et al.* [1], questo evento rappresenta una pietra miliare negli studi di antropologia molecolare, genetica medica e genetica forense, essendo da quel momento in avanti utilizzata come riferimento indispensabile per l'osservazione dei polimorfismi presenti sulla catena leggera del genoma mitocondriale analizzato. Per tale ragione viene definita appunto: sequenza di riferimento di Cambridge (RCS). Nel corso degli anni, alcune discrepanze riportate in letteratura [115] hanno spinto nel 1999 Andrews *et.al* a rianalizzare la placenta umana di origine Europea utilizzata per il primo sequenziamento, confermando la presenza di undici errori e di sette rari polimorfismi a carico della CRS (Tab.3.1). Uno degli errori osservati più importanti, riguarda la perdita di una citosina in posizione 3106 che avrebbe quindi ridefinito la corretta lunghezza della sequenza di riferimento. Tuttavia l'accorciamento della sequenza avrebbe apportato una maggiore confusione interpretativa, in particolare per gli studi già presenti in letteratura. Andrews *et.al* decisero quindi di mantenere la precedente numerazione inserendo una

delezione in posizione 3107. Fortunatamente nessun errore viene rilevato a all'interno della regione di controllo della CRS. Queste ambiguità possono essere giustificate a causa delle prime rudimentali procedure di sequenziamento utilizzate, colmate all'epoca da Anderson *et al.*, inserendo sequenze di cellule tumorali immortalizzate HeLa e DNA mitocondriale bovino. La sequenza corretta viene quindi definita col nuovo termine di *revised Cambridge Reference Sequence* (rCRS) ed accettata come nuovo riferimento per le comparazioni, e sulla base della citosina presente in posizione 14766 e 7028 viene assegnata all'aplotipo H.

Posizione	Sequenza precedente	Sequenza rianalizzata	Note
263	A	A	polimorfismo raro
311-315	CCCCC	CCCCC	polimorfismo raro
750	A	A	polimorfismo raro
1438	A	A	polimorfismo raro
316-3107	CC	C	errore
3423	G	T	errore
4769	A	A	polimorfismo raro
4985	G	A	errore
8860	A	A	polimorfismo raro
9559	G	C	errore
11335	T	C	errore
13702	G	C	errore
14199	G	T	errore
14272	G	C	errore (bovino)
14365	G	C	errore (bovino)
14368	G	C	errore
14766	T	C	Errore (HeLa)
15326	A	A	polimorfismo raro

TABELLA 3. 1. Discrepanze osservate nel confronto con la CRS (Anderson *et al.* 1981) e rCRS (Andrews *et al.* 1999).

Capitolo 4: Scopo dello Studio

Nell'ambito di una collaborazione tra il Dipartimento di Chimica Biologica e Biotecnologie dell'Università di Perugia ed il laboratorio di genetica forense dell'Università di Medicina di Innsbruck, è stato condotto uno studio di genetica delle popolazioni, tipizzando il DNA mitocondriale appartenente a cento-cinquantuno individui di origine Mongola, attraverso l'impiego del *Next Generation Sequencing*. In questo contesto, poiché l'NGS è già stato adottato con entusiasmo in diversi settori scientifici che spaziano dall'antropologia molecolare alla genetica medica, si inserisce uno studio parallelo condotto dal laboratorio di genetica forense dell'Università di Medicina di Innsbruck assieme al laboratorio di genetica forense del dipartimento di Scienze Mediche e Chirurgiche dell'Università di Bologna, con due finalità: la prima, valutare il *Next Generation Sequencing* nell'analisi genetico-forense per una sua futura validazione, anche alla luce di un pannello di recentemente sviluppo specifico per l'analisi dell'intera sequenza del genoma mitocondriale. Ciò consentirebbe di superare le limitazioni imposte dal metodo di Sanger, ad oggi quello universalmente accettato in sede giudiziaria, caratterizzate da: un basso nel numero di campioni processati per volta, difficoltà nell'analisi di estese regioni genomiche, oltre all'aumento dei costi e delle tempistiche. A tal proposito lo studio valuta l'affidabilità analitica dell'NGS in relazione ad alcune problematiche interpretative da sempre presenti nel mondo della genetica forense rappresentate dalle varianti di lunghezza in prossimità dei tratti di omopolimeri di citosina (*C-stretches*) spesso responsabili di artefatti, e dalle eteroplasmie puntiformi la cui corretta interpretazione può fornire un dettaglio estremamente importante nell'associazione tra due individui all'interno di una discendenza. La seconda finalità è rappresentata dalla valutazione del potere discriminativo della *Next Generation Sequencing* comparando l'intera sequenza del mitogenoma alla sequenza della sola regione non codificante di controllo, attualmente l'unica considerata nella routine forense anche per ragioni etico-giuridiche. A questo fine verranno calcolati i parametri

forensi che includono gli indici standard di diversità genetica e la *Random Match Probability* intesa come indice della variabilità all'interno del database analizzato.

I risultati ottenuti consentiranno di ampliare il database di popolazione EMPOP (<http://empop.online/>) utilizzato a scopo forense nel calcolo statistico di probabilità per determinare l'associazione tra due profili genetici. Inoltre, la comparazione tra NGS e metodo di Sanger aggiungerà un ulteriore contributo verso una futura validazione di questo sistema anche nelle analisi forensi di routine.

Capitolo 5: Materiali e Metodi

5.1 Campioni, provenienza ed estrazione

Campioni di DNA da 151 individui non imparentati tra loro, provenienti da 21 province della Mongolia e di genealogia certificata tramite pedigree, sono stati forniti dal Dipartimento di Chimica, Biologia e Biotecnologie dell'Università di Perugia per l'analisi dell'intero genoma mitocondriale tramite Next Generation Sequencing, in collaborazione con il laboratorio di genetica forense dell'Università di Medicina di Innsbruck . I campioni provengono da un set più ampio di 3000 individui, frutto di un precedente progetto mirato all'analisi delle sole regioni ipervariabili sulla CR del mitogenoma per lo studio della variabilità genetica della Mongolia moderna, che ha coinvolto la Sorenson Molecular Genealogy Foundation, l'Università Nazionale della Mongolia ed il Dipartimento di Chimica, Biologia e Biotecnologie dell'Università di Perugia. Il DNA è stato estratto tramite Wizard® Genomic DNA Purification Kit (Promega, Madison, WI) da campioni di saliva.

5.2 Preparazione della libreria genomica

Data la dimensione totale del genoma mitocondriale di ~16Kb il normale sequenziamento di Sanger risulta inadeguato per l'analisi di sequenze così estese. Ion PGM System™ (Thermo Fisher Scientific, Waltham, MA, USA) , appartenente alla recente tecnologia Next Generation Sequencing, offre invece un sistema ad alta resa ed efficienza per la produzione di un elevato numero di dati in breve tempo. Tramite il pannello sperimentale Ampliseq Precision ID (Thermo Fisher Scientific, Waltham, MA, USA) specificatamente sviluppato per il mitogenoma e tutt'ora in fase di studio

(Prof. Walther Parson personal communicator), è possibile infatti svolgere l'analisi dell'intera sequenza con NGS. Per analizzare la sequenza del DNA mitocondriale è necessaria la preparazione di una libreria genomica per ciascun campione, amplificando almeno 1ng di mtDNA estratto, attraverso due pool di *primers* specifici (dati non mostrati) per ottenere una serie di frammenti della dimensione di ~175bp rappresentativi dell'intero mitogenoma. Per evitare un'eventuale perdita di informazioni durante il sequenziamento, tutti i frammenti vengono creati in modo che risultino tra loro sovrapposti alle estremità. Viene quindi allestita la mix di reazione in un volume finale di 20µl in una piastra da 96 pozzetti secondo protocollo:

Componenti	Volume
5X Ion Ampliseq™ HI-FI Mix	4µl
Primer Pool 1 / Primer Pool 2	10µl
>1ng gDNA / mtDNA	Xµl
ddH2O nuclease free	(6-X)µl

Ed effettuata tramite 2720 Thermal Cycler (Applied Biosystem, Foster City, CA) Il profilo termico della reazione di amplificazione consiste in:

Step	Temperatura	Tempo	
Hold	attivazione della polimerasi	99°C	2 min
21 cicli	deanturazione	99°C	15 sec
	annealing / estensione	60°C	
Hold	-	10°C	4 min

Nella fase seguente la libreria genomica di DNA mitocondriale viene allestita accorpando i prodotti amplificati in una nuova piastra, trasferendo in ciascun pozzetto 10 µl di ampliconi corrispondenti ai *primers* Pool 1 e Pool 2 dei rispettivi campioni, in un volume finale di 20 µl. I *primers* incorporati vengono parzialmente digeriti enzimaticamente con l'aggiunta in ciascun pozzetto di 2 µl di FuPa, per rendere le estremità degli ampliconi *blunt*, incubando secondo protocollo:

Temperatura	Tempo
50°C	10 min
55°C	10 min
60°C	20 min
10°C	Hold

Per sequenziare più campioni nello stesso momento, la preparazione prevede per ciascun frammento della libreria, il ligaggio alle estremità 5' e 3' di specifiche sequenze *barcode*. E' necessario allestire una mix di Ion P1 adapter e Ion Xpress™ Barcode X, dove ogni *barcode* è diluito 1:4 combinando i volumi secondo quanto riportato dal protocollo nella tabella sottostante.

Barcode mix	Volume
Ion P1 adapter	2µl
Ion Xpress™ Barcode X	2µl
H2O nuclease free	4µl
Totale	8µl

Per ogni campione si aggiungono 2µl di *barcode* mix e 4µl di Switch Solution, ottenendo un volume finale di 28 µl per pozzetto, a cui si aggiungono infine 2µl di enzima DNA Ligasi, attivando la reazione di ligaggio alle seguenti condizioni:

Temperatura	Tempo
22°C	30 min
72°C	10 min
10°C	Hold

I frammenti vengono quindi purificati da eventuali contaminanti di PCR quali: sali, enzimi, primers e dNTPs non incorporati. Il passaggio richiede l'aggiunta al campione di 45 µl di Agencourt® AMPure® XP contenente una soluzione di biglie para-magnetiche capaci di legare reversibilmente tutti i frammenti amplificati. Segue la centrifugazione della piastra per 5 minuti a 2000rpm per far precipitare le biglie, dopo averla posizionata su un apposito rack magnetico è quindi possibile

rimuovere il sovranatante. L'ultimo passaggio della purificazione consiste in due sessioni di lavaggio delle biglie, aggiungendo per ogni campione 150 µl di EtOH 70% e rimuovendo il sovrinatante. Una volta eliminato l'etanolo residuo, si procede al rilascio del DNA complessato alle biglie aggiungendo al pellet 50µ di Buffer Low TE, un chelante per gli ioni metallici, centrifugando infine la piastra per 5 minuti a 2000rpm.

Recuperato il sovrinatante si procede alla quantificazione delle librerie genomiche tramite real-time PCR allestendo una nuova piastra in cui disporre in duplicato, 9µl da aliquote di campioni diluiti rispettivamente 1:100 ed 1:5000, controllo negativo (NTC) e 9µl di standard interno E. Coli DH10B Ion Control Library diluito serialmente secondo tabella:

Standard	E. Coli Control Library	H ₂ O	Conc [pM]
1	5 µl non diluito	45 µl	6,8
2	5 µl Std.1	45 µl	0,68
3	5 µl Std.2	45 µl	0,068
4	5 µl Std.3	45 µl	0,0068
5	5 µl Std.4	45 µl	0,00068

Agli standard ed ai campioni si aggiungono 11µl di qPCR reaction mix preparata secondo protocollo:

qPCR reaction mix	µl
Ion Library TaqMan® qPCR Mix	10
Ion Library TaqMan® Quantitation Assay, 20X	1

cui segue la real-time PCR con Applied Biosystem 7500 Real-Time PCR System (Applied Biosystem, Foster City, CA) secondo profilo termico:

Step	Temperatura	Tempo
Hold	50°C	2 min
Hold	95°C	20 sec
40 cicli	95°C	3 sec
	60 °C	32 sec

Per iniziare l'analisi del mitogenoma attraverso Ion PGM System™, è necessario che le librerie genomiche dei campioni in analisi, abbiano la stessa concentrazione di 20pM. I campioni vengono quindi diluiti con nuclease free H₂O in base alla loro concentrazione media, utilizzando la seguente formula:

$$\text{Template Dilution Factor} = [\text{Avg Conc of Undiluted Lib}]/20 \text{ pM}$$

Dai campioni diluiti, si prelevano rispettivamente aliquote dello stesso volume che verranno miscelate assieme in un'unica mix complessiva di tutte le librerie mtDNA prodotte. Da questa mix, vengono infine prelevati 10µl e uniti a 15µl di nuclease free H₂O ottenendo un soluzione finale di 25µl necessaria per il primo passaggio della preparazione dei *template* nell'analisi delle sequenze.

5.3 Sequenziamento con Ion PGM System™

Una volta create le librerie, l'analisi del mitogenoma con PGM è preceduta da due fasi preparatorie di emulsione ed arricchimento. L'emulsione prevede l'allestimento della soluzione di amplificazione come mostrato nella seguente tabella sottostante:

Reagente	Volume
Ion PGM™ Hi-Q™ Reagent Mix	800µl
H ₂ O Nuclease Free	25 µl
Ion PGM™ Hi-Q™ Enzyme Mix	50µl
Mix librerie mtDNA diluite	25µl
Ion PGM™ Hi-Q™ ISPs	100µl

che viene quindi caricata assieme ad 1.7 ml di Ion OneTouch™ Reaction Oil, nel filtro di reazione all'interno di Ion OneTouch™ 2 (Thermo Fisher Scientific, Waltham, MA, USA), in cui avviene l'amplificazione clonale dei frammenti di ciascuna delle librerie, complessati alle biglie Ion PGM™ Hi-Q™ ISPs. Segue la purificazione degli ampliconi, che consiste nella rimozione di 100µl di sovrantante; nella risospensione del pellet in 500µl di Ion OneTouch™ Wash Solution , ed infine 1.5 min di centrifugazione a 11000 rpm per rimuovere la Wash Solution fino ad ottenere un volume finale di 100µl di soluzione.

La fase preparatoria all'arricchimento richiede il trasferimento dei 100µl di soluzione purificata in una strip da otto pozzetti: assieme a 130µl di una miscela formata da Dynabeads® MyOne™ Streptavidin C1 e MyOne™ Beads Wash Solution; 300 µl Ion OneTouch™ Wash Solution; 300 µl di Melt-off solution preparata secondo protocollo:

Componente	Volume
Tween® Solution	280µl
1 M NaOH	40µl

Il processo d'arricchimento verrà infine svolto dallo strumento Ion OneTouch™ ES (Thermo Fisher Scientific, Waltham, MA, USA) per scremare dalla soluzione i frammenti amplificati non complessati alle biglie Ion PGM™ Hi-Q™ ISPs.

Prima dell'analisi del mitogenoma con Ion PGM System™, è necessaria una fase preparatoria di lavaggio dello strumento utilizzando una soluzione a base di Cloro cui segue risciacquo con H₂O 18MΩ (Elga Purelab Flex 3 Water Purification, Veolia Water System, Austria), ed un Chip Check Test durante l'inizializzazione dell' Ion Torrent utilizzando un Chip vuoto per verificare il corretto funzionamento del sistema. La soluzione purificata ed arricchita viene infine caricata su Ion 318™ Chip per il sequenziamento NGS dell'intero mitogenoma.

5.4 *Analisi ed interpretazione delle varianti tramite NGS*

Tutte le sequenze ottenute con NGS sono allineate nel Torrent Server utilizzando la sequenza di riferimento di Cambridge (rCRS), ed analizzate dal software Ion Torrent Software Suite V.4.4.3; V.4.6 e V.5.0.5. utilizzando il *plug-in* variant caller V.5.2.1.38 tramite allineamento ottimizzato TMAP Smith-Waterman [116]. Il risultato è una una lista delle differenze individuate tra il campione analizzato e la rCRS, che viene confrontata assieme alla corrispondente rappresentazione grafica fornita dal software IGV V.2.3 (Broad Institute, Cambridge, MA, USA) [117] (Fig.5.1) attraverso i file BAM (binary alignment map) e BAI (binary alignment index) prodotti dal Torrent Server.

In accordo con le linee guida Forensi redatte dalla Società Internazionale di Genetica Forense (ISFG) [97] per la corretta interpretazione delle sequenze del mtDNA, le inserzioni o le delezioni di basi osservate sia nei tratti omopolimerici sia in altre sequenze ripetute del genoma, vengono allineate in posizione relativa al 3'. La coesistenza di mtDNA multipli all'interno di una cellula, indicata col termine di eteroplasmia, viene valutata adottando una soglia di *coverage* totale del 15% nel caso di eteroplasmia puntiforme (PHP), mentre le eteroplasmie di lunghezza (LHP) osservate in prossimità dei tratti di policitosina quali: 309.1C; 315.1C; 515-522 indels e 16193.1C vengono riportate ma non incluse nell'analisi filogenetica essendo comunemente presenti all'interno del genoma mitocondriale, annotando come richiesto dalle stesse linee guida la variante dominante. Per tutte le eteroplasmie osservate viene inoltre svolto un ulteriore controllo, che consiste nel confronto casuale della posizione considerata con cinque campioni appartenenti alla stessa sessione e cinque campioni di sessioni diverse.

L'insieme delle posizioni mutate che costituiscono l'aplotipo viene utilizzato per l'affiliazione dell'aplogruppo d'appartenenza, utilizzando *PhyloTree*, versione Build 17, un accurato albero filogenetico costruito su 5400 aplogruppi indicativi delle posizioni presenti sull'intero mitogenoma

[118] e confermato col database online EMPOP V.3/R11 (<http://empop.online/>) che si basa sull'impiego di due algoritmi quali: SAM, un algoritmo di ricerca che confronta le posizioni dell'aplotipo o le sequenze FASTA assieme alle sequenze presenti all'interno del database, tramite un approccio che non utilizza il classico allineamento, reso impossibile per gli eventi di inserzione e delezione che hanno luogo sulla molecola. Inoltre, nell'ultima versione di EMPOP, l'algoritmo è stato ulteriormente implementato per riflettere al meglio la filogenesi del mitogenoma. Il secondo algoritmo EMMA permette l'assegnazione corretto aplogruppo confrontando un totale di 14.990 genomi di DNA mitocondriale presenti all'interno del database di sequenze GenBank®, assieme a 3925 aplotipi virtuali ottenuti da Phylotree e 19.171 sequenze della regione di controllo, utilizzando il massimo approccio probabilistico che tiene in considerazione le fluttuazioni dei tassi di mutazione rappresentato dalla formula descritta da Rock et al.[96]:

$$r_{\alpha\beta} = \frac{\sum_{\gamma} \min(n(\alpha, \gamma), n(\beta, \gamma))}{\sum_{\gamma} n(\gamma)}$$

In cui α e β rappresentano le basi A,T,G,C e dove α è diverso da β , la variabile γ rappresenta i clusters degli aplogruppi della CR (CR-HG) in cui α o β sono dominanti, $n(\gamma)$ indica il numero totale di campioni in CR-HG γ . L'algoritmo compara quindi il profilo inserito con ciascuno dei database, assegnando le differenze osservate attraverso la corretta stima del *cost* che è definita come la divergenza tra due profili.

5.5 Sequenziamento di Sanger

Per valutare la robustezza analitica del *Next Generation Sequencing*, la regione di controllo e le eteroplasmie puntiformi individuate nella regione codificante, sono state comparate attraverso il sequenziamento di Sanger (STS), che attualmente rappresenta il *gold standard* nel campo forense. Il metodo prevede l'amplificazione di 2 μ l di DNA assieme a 4 controlli negativi, in 18 μ l di PCR

master-mix contenente: 1 U AmpliTaq Gold polymerase, 1U di PCR Reaction Buffer, 200 μ M singoli dNTP (Thermo Fisher Scientific, Waltham, MA, USA) e 0.5 μ l di ciascun primer (Tab.5.1) (L15900 e H00599) per la reazione di amplificazione della regione di controllo.

Primer	Sequenza
L15900	5'TCA AAG CTT ACA CCA GTC TTG TAA ACC3'
H00599	5'TTG AGG AGG TAA GCT ACA TA 3'

TABELLA 5. 1. Sequenze dei *primers* utilizzati per la prima amplificazione della regione di controllo.

la reazione di PCR viene svolta tramite Gene Amp 9700 PCR System (Applied Biosystem, Foster City, CA) seguendo il profilo termico rappresentato in tabella:

	Step	Temperatura	Tempo
Hold	attivazione della polimerasi	95°C	11 min
	deanturazione	95°C	15 sec
35 cicli	annealing	56°C	30 sec
	estensione	72°C	90 sec
Final extension		72°C	10 min

Gli ampliconi sono purificati con 8 μ l di Exo-SAP-IT (USB, Cleveland, OH) a 37°C per 15 minuti e 80°C per 15 minuti, per rimuovere i *primers* e dNTPs non incorporati.

Due microlitri di prodotti purificati vengono quindi dispensati all'interno della master-mix di sequenziamento di Sanger, contenente 2 μ l Big Dye Terminator V.1.1 Cycle sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA) in un volume finale di 8 μ l, contenente:

Componente	Volume
Big Dye Terminator V.1.1 RR mix	2 μ l
Big Dye Terminator V.1.1 Sequencing Buffer	2 μ l
Primer di sequenziamento	16pmol
ddH ₂ O	a volume (a 8 μ l)

Il sequenziamento viene svolto ai seguenti cicli di temperature che comprendono: una prima denaturazione iniziale a 95°C per 1 minuto e 25 cicli a 95°C per 10 secondi, 50°C per 5 secondi e 60°C per 4 minuti. Ogni amplicone è stato sequenziato in direzione *forward* utilizzando i *primers* L15971, L15989, L16268, L00015, L00029, L00314 e L00361, ed in direzione *reverse* utilizzando i *primers* H00016, H00159 e H00484 (Fig.5.2).

I prodotti di sequenziamento sono infine purificati secondo protocollo con Sephadex G-50 Fine (Amersham, Buckinghamshire, UK), diluiti aggiungendo 10µl di acqua distillata e separati tramite elettroforesi capillare al sequenziatore ABI3100 (Applied Biosystem, Foster City, CA) utilizzando per la corsa elettroforetica polimero dimetil-poliacrilamide POP6 [119], con capillare di 36cm di lunghezza. Le sequenze di ciascun campione ottenute sotto forma di elettroferogrammi, vengono infine interpretati tramite il software Sequencer V.5.4 (Gene Codes, Ann Arbor, MI, USA).

Primer	Sequenza
H00016	5' TGA TAG ACC TGT GAT CCA TCG TGA 3'
H00159	5' AAA TAA TAG GAT GAG GCA GGA ATC 3'
H00484	5' TGA GAT TAG TAG TAT GGG AG 3'
L00015	5' CAC CCT ATT AAC CAC TCA CG 3'
L00029	5' CTC ACG GGA GCT CTC CAT GC 3'
L00314	5' CCG CTT CTG GCC ACA GCA CT 3'
L00361	5' ACA AAG AAC CCT AAC AAC AGC 3'
L15971	5' TTA ACT CCA CCA TTA GCA CC 3'
L15989	5' CCC AAA GCT AAG ATT CTA AT 3'
L16268	5' CAC TAG GAT ACC AAC AAA CC 3'

TABELLA 5. 2. Sequenze dei *primers* utilizzati nel sequenziamento della regione di controllo.

5.6 *Calcolo dei parametri genetico-forensi*

L'analisi variabilità genetica della popolazione Mongola sotto il profilo forense, viene svolta calcolando il valore di *Haplotype Diversity* ed il numero di aplotipi unici generati sia per l'intero mitogenoma sia per le regioni ipervariabili HVS I e II, assieme alla stima del valore medio di differenze tra gli aplotipi, inteso come *mean pairwise difference* (MPD), tramite il software ARLEQUIN v.3.5.1.2 [120]. La probabilità che due individui scelti a caso nella popolazione analizzata abbiano lo stesso aplotipo, intesa come *Random Match Probability* (RMP), viene calcolata per l'intero genoma mitocondriale e per le regioni ipervariabili HVS I e II, attraverso la somma delle frequenze elevate al quadrato di ciascun aplotipo per rappresentare il grado di differenziamento all'interno della popolazione. Le varianti di lunghezza introdotte dalle *C-stretches* in posizione: 310; 573; 16193; 16199 presenti nelle regioni ipervariabili HVS I, HVS II e HVS III, sono state escluse per la caratterizzazione dell'aplotipo per il calcolo della RMP.

Capitolo 6: Risultati

6.1 *Analisi filogenetica e molecolare del genoma mitocondriale*

Dal sequenziamento dell'intera sequenza del genoma mitocondriale attraverso Ion PGM System™, utilizzando il pannello sperimentale Ampliseq Precision ID (Thermo Fisher Scientific) per 151 campioni di DNA proveniente da individui originari della Mongolia, è stato ottenuto un totale di 146 profili genetici completi, mentre cinque profili parziali hanno riportato almeno un *drop-out* tra le posizioni: 10831-10843; 8487-8507; 2129-2131,8475-8507, 8716-8784,14755-14799; 2101-2136, 8707-8750; 142-234, 8499-8507, 8674-8762. Sebbene per otto campioni, è stata osservata una probabile contaminazione, per tutti è stato comunque possibile determinare l'aplotipo corrispondente. L'inferenza dell'aplogruppo svolta con Phylotree versione Build 17 e confermata con EMPOP V.3/R11, ha consentito di individuare all'interno del set di campioni 115 linee (Tabella A.1). Poiché l'analisi dell'intero genoma mitocondriale ha permesso di distinguere ulteriori sotto-linee, la maggior parte è stata raccolta all'interno dei principali macro-aplogruppi mostrati in Tabella 6.1 con le relative frequenze riportate.

All'interno del campione analizzato è stata osservata una maggioranza di linee appartenenti ai macro-aplogruppi H (27%), U (25%), T (8%) e J (7.3%), cui seguono gli aplogruppi HV (6.6%) prevalentemente diffuse in Europa ed Asia meridionale, D4 (6.6%) frequente in Asia orientale, X2 (4%) in Europa, K e W (2%) rispettivamente in Europa ed Asia centrale. Allo stesso modo, anche gli aplogruppi individuati singolarmente nel campione sono tutti riconducibili all'Asia orientale, meridionale ed Europa.

Aplogruppo	n	(%)	Aplogruppo	n	(%)
B4	7	4,6	M10a1+16129	1	0,7
C1a	1	0,7	M7b1b	1	0,7
D4	10	6,6	R11b	1	0,7
F	3	2,0	R1a1	1	0,7
H	41	27,2	R2	1	0,7
HV	10	6,6	T	12	7,9
I1c1	1	0,7	U	38	25,2
J	11	7,3	W	3	2,0
K	3	2,0	X2	6	4,0

TABELLA 6.1. Macroaplogruppi osservati all'interno del campione di 151 individui rappresentativi della popolazione Mongola. I sotto-aplogruppi unici sono stati riportati come tali.

Per tutti i campioni, l'analisi dell'intera sequenza ha individuato 748 siti polimorfici in 16569 loci, che ha originato un totale di 142 aplotipi, di cui 133 unici e 9 condivisi. Dalla sola regione di controllo invece, 175 siti polimorfici osservati in 1132 loci hanno prodotto 133 aplotipi, di cui 111 unici e 18 condivisi. In entrambi i casi, le transizioni di base rappresentano il 94% delle mutazioni totali mentre le trasversioni il 6%. Come mostrato in Tabella 6.2, l'esame della variabilità basata sulle differenze tra i singoli aplotipi, ha riportato un maggior valore sia di *Haplotype Diversity* sia di *Mean Pairwise Differences* per l'intero genoma mitocondriale rispetto alla sola regione di controllo, mentre la probabilità di osservare lo stesso aplotipo all'interno del *dataset* ottenuto dalla popolazione Mongola, intesa come *Random Match Probability*, risulta 1:119 per la CR e 1:135 per tutta la sequenza.

	CR+CodR	CR
Aplotipi osservati	142	133
Aplotipi unici	133	111
Aplotipi condivisi	9	18
Haplotype Diversity	0.999	0.998
Mean Pairwise Differences	34.340	11.109
RMP ($p=\Sigma x^2$)	$7,397 \times 10^{-3}$	$8,358 \times 10^{-3}$

TABELLA 6.2 indici standard di diversità riportati, per l'intera sequenza del mitogenoma e per la sola regione di Controllo

6.2 *Analisi del genoma mitocondriale con Next Generation Sequencing*

A partire da un 1ng di DNA mitocondriale, Ion PGM System™ ha sequenziato una media per campione di 2.787.308 frammenti (*reads*) il cui valore di *usable reads* corrispondente alla quantità netta utilizzabile per le analisi rappresentata in percentuale, risulta per la maggior parte delle sessioni analitiche inferiore al 50%. Allo stesso modo, l'amplificazione clonale di tutti i frammenti delle librerie genomiche attraverso *emulsion* PCR, ha prodotto una resa superiore al 50% solamente per sei delle quattordici sessioni analitiche caricate su Ion 318™ Chip (Tab.6.3). Il *coverage* medio ottenuto per ciascuna posizione analizzata in tutti i genomi, è compreso in un intervallo tra 96 e 13484, registrando nella quasi totalità dei campioni, i valori più bassi per le posizioni 8674-8764. Delle 4372 varianti SNPs identificate col confronto della sequenza di riferimento di Cambridge tramite il *plug-in variant caller* V.5.2.1.38, il 7% è rappresentato da artefatti analitici localizzati nelle posizioni 249, 4824, 6641 e 8373 (Fig.A.1; A.2; A.3; A.4), in proporzioni variabili nella maggior parte degli allineamenti analizzati. Allo stesso modo, segnali analitici non corrispondenti a reali eteroplasmie puntiformi per le posizioni: 751,752,756,769,770,762,786 e 790, sono state osservate durante il confronto diretto per tutti i 151 allineamenti (Fig.A.5), sebbene non siano state identificate dal *plug-in*.

Nonostante in letteratura non venga fatta una reale distinzione tra eteroplasmie di lunghezza (LHP) e varianti di lunghezza, in questo studio utilizzeremo LHP per definire tutte le inserzioni o delezioni di nucleotidi osservate in campioni singoli, mentre per varianti di lunghezza intendiamo tutte quelle inserzioni o delezioni osservate in un ampio numero di campioni per le stesse posizioni del genoma mitocondriale. Almeno una variante di lunghezza causata da un inserzione o delezione di base è presente nel 98% di tutti i genomi analizzati. Nello specifico: la maggior parte è stata osservata nella CR in prossimità delle regioni ipervariabili HVS I, II e III (315.1C, 523-524 *indel*, 16193.1C) a livello delle C-stretch ed in sequenze ripetute di nucleotidi.

Sessione	Reads	Usable Reads (%)	Amplificazione monoclonale (%)	Amplificazione policlonale (%)
Run 1	4029732	65	73	27
Run 2	4672586	68	75	25
Run 3	3551754	42	52	48
Run 4	327162	44	53	47
Run 5	3239131	47	58	42
Run 6	2464444	26	37	63
Run 7	2926615	39	54	46
Run 8	3311919	35	44	56
Run 9	1906007	25	38	62
Run 10	2588618	33	44	56
Run 11	2449168	35	45	55
Run 12	2504261	27	38	62
Run 13	2688476	36	46	54
Run 14	2362441	27	38	62

TABELLA 6.3. numero medio di frammenti prodotti (*reads*) per sessione analitica e la quantità netta utilizzabile in relazione all'amplificazione monoclonale e policlonale delle librerie

Al contrario, solamente il 5% delle duecento-cinquantadue varianti di lunghezza osservate all'interno del nostro set di campioni, sono invece presenti nella regione codificante (8281-8289DEL, 15944DEL, 2405.1C, 5752DEL, 965.1C, 8276.1C). Nessuna LHP è stata invece individuata nel *set* analizzato. In cinque campioni Ion PGM System™ non è in grado di fornire con esattezza il numero di basi di citosina, inserite nella regione codificante, per la variante di lunghezza 8278.XC presente in diversi sotto-aplogruppi appartenenti alle linee D, HV, L,M, R,S, U5.

Uno degli aspetti più importanti nell'ambito dello studio presentato, riguarda la valutazione delle eteroplasmie puntiformi (PHP). Per la loro discriminazione con Ion PGM System™ è stata adottata una soglia di rilevazione pari al 15% della componente minoritaria nella miscela (*Minor Allele Frequency*), che ha identificato 25 posizioni eteroplasmiche totali. Solamente in un individuo è stata rilevata la presenza simultanea di due PHP all'interno del profilo, mentre nessuna è stata trovata condivisa. Data l'elevata sensibilità strumentale, per eliminare possibili artefatti introdotti dal rumore di fondo, tutte le PHP individuate sia attraverso il *plug-in variant caller*, sia attraverso l'analisi diretta della sequenza, sono state confrontate con cinque allineamenti scelti casualmente

all'interno della stessa sessione analitica, e cinque provenienti da sessioni diverse. Un insieme di sei posizioni presumibilmente eteroplasmiche: 189R, 214R, 7521R, 8152R, 8373R e 8943Y, rispettivamente per i campioni Mg2_E01; Mg_5; Mg1_B03; Mg1_D04; Mg1_F09 e Mg2_F09, sono state quindi confrontate e sequenziate tramite metodo di Sanger che ha quindi permesso di discriminare come artefatti due posizioni, rispettivamente: 8373R in Mg1_F09 e 8943Y in Mg2_F09 (Fig.A.6a/b; Fig.A.7a/b; Fig.A.8a/b; Fig.A.9a/b; Fig.A.10a/b; Fig.A.11a/b). Le 23 PHP certe (Tab.6.4), rappresentano il 15% di tutti gli individui analizzati e sono disposte nel genoma mitocondriale in maniera omogenea. Undici posizioni sono infatti localizzate sulla regione di controllo, e 10 sulla regione codificante, quattro delle posizioni osservate: 152, 189, 214, 16278 rientrano all'interno del *set* di mutazioni *hotspot* descritto in letteratura [62;70;71]. Infine, la maggior parte delle eteroplasmie consiste in transizioni a carico di citosina a timina mentre solo un individuo risulta caratterizzato da una trasversione da citosina ad adenina

Il confronto tra il sistema NGS ed il metodo di Sanger (STS) nella regione di controllo, ha evidenziato alcune discordanze nell'interpretazione delle varianti di lunghezza e delle sostituzioni di basi per le regioni ipervariabili HVS I e II (Tab.6.5). Nel primo caso, per 78 genomi le inserzioni di citosina individuate con Sanger in posizione 309.1C non sono invece rilevate dal *variant caller* di Ion PGM System™ così come le venti inserzioni dinucleotidiche AC in posizione 563 riportate nell'aplotipo dei campioni MG_3 e Mg2_A06 con Sanger. Allo stesso modo l'inserzione 16192.1T individuata tramite STS nei campioni Mg1_E01 e Mg_E07, viene interpretata come mutazione C16192T in NGS. Al contrario, l'inserzione 16193.1C individuata dal sequenziamento NGS per i campioni MG1_F06 ed MG2_F05, non risulta presente con STS. Un'ulteriore discrepanza, riguarda l'assegnazione del preciso numero di basi inserite all'interno delle varianti di lunghezza in posizione 16193 per i campioni Mg2_D04 e Mg1_2HG02.

Per quel che riguarda le sostituzioni di base, è interessante notare che: per il campione MG2_C04, le mutazioni riportate in Sanger G118C, T119C, G143A, 146Y, A153G, A189G, C194T, T195C, T196C, T204C e G207A, sono tutte definite in MPS come eteroplasmie puntiformi (PHP), così come le posizioni 239C 16362C 16278T 16311C nel campione MG2_C06. Sempre all'interno dello stesso campione inoltre, NGS individua rispetto a Sanger un'ulteriore sostituzione G16274A.

Campione	Posizione	Eteroplasmia	Rcrs	Variante	Regione
Mg1_A02	151	Y	C (67%)	T (33%)	CR
Mg2_F09	152	Y	T (33%)	C (67%)	CR
Mg2_E01	189	R	A (84%)	G (16%)	CR
MG5	214	R	A (85%)	G (15%)	CR
Mg1_A12	246	Y	T (43%)	C (57%)	CR
Mg2_A04	267	Y	T (57%)	C (43%)	CR
Mg2_E04	4335	Y	C (26%)	T (74%)	codR
Mg1_F03	4387	Y	C (58%)	T (42%)	codR
Mg2_B05	5746	R	G (35%)	A (65%)	codR
Mg1-2H_F02	7028	Y	C (80%)	T (20%)	codR
Mg1_B03	7521	R	G (33%)	A (67%)	codR
Mg1_D04	8152	R	G (76%)	A (24%)	codR
Mg1_B04	8463	R	A (15%)	G (85%)	codR
Mg2_E05	8778	M	C (29%)	A (71%)	codR
Mg1_D06	11674	Y	C(68%)	T(32%)	codR
Mg1_D06	11947	R	A(74%)	G(26%)	codR
Mg2_C07	13759	R	G (53%)	A (47%)	codR
Mg1_B05	15727	Y	C (47%)	T (53%)	codR
Mg1_B08	16093	Y	T (21%)	C (79%)	CR
Mg1_C02	16126	Y	T (67%)	C (33%)	CR
Mg1_D05	16172	Y	T (53%)	C (47%)	CR
Mg1_C12	16278	Y	C (77%)	T (23%)	CR
Mg1_A05	16381	Y	T (51%)	C (49%)	CR

TABELLA 6.4. Eteroplasmie puntiformi (PHP) riportate secondo la corretta terminologia IUPAC intesa come: Y = miscela di C/T; R = miscela di A/G; M = miscela di G/A; assieme alle rispettive percentuali di varianti maggiore e minore, posizione e regione

Campioni (n)	Posizione	NGS	STS
78	309		.1C -.3C
2	563		.1C -.20C
2	16192	C/T	.1T
2	19193	.1C	
2	16193	.1C	.2C
1	118	S	C
1	119	Y	C
1	143	R	A
1	146	Y	Y
1	153	R	G
1	189	R	G
1	194	Y	T
1	195	C	C
1	196	Y	C
1	204	Y	C
1	207	R	A
1	239	Y	T/C
1	16278	Y	C/T
1	16311	Y	T/C
1	16362	Y	T/C
1	16482	R	A/G
1	16274	G/A	
1	8373	R	
1	8943	Y	

TABELLA 6.5. Discordanze riportate nel confronto tra il metodo di Sanger ed il sequenziamento NGS. Le terminologie IUPAC utilizzate in tabella si riferiscono a: Y = miscela di C/T; R = miscela di A/G.

Capitolo 7: Discussione

L'analisi di 155 campioni di DNA mitocondriale proveniente da individui originari della Mongolia, ha evidenziato un'ampia variabilità di linee e sotto-linee filogenetiche non esclusive del continente Asiatico, che indicano una notevole commistione con l'Europa occidentale. Questo risultato, assieme alla prevalenza degli aplogruppi Europei H, U e J osservati, è coerente con un precedente studio svolto sulla popolazione Mongola che ha interessato la regione di controllo di 201 mitogenomi [121]. Le cause di questa mescolanza tra le due aree può essere ricercata sia nella presenza della via della seta che attraversando l'Asia centrale [122] può aver fornito un importante contributo all'ingresso degli aplogruppi Europei in Asia, sia nell'Impero Mongolo (1206-1368 dC) fondato da Gengis Khan (1162-1227dC), che aveva incluso diverse popolazioni Eurasiatiche per un estensione di 3.5 milioni di km², portando a una consistente mescolanza delle linee Asiatiche con quelle Europee [121].

In questi ultimi anni, l'analisi combinata dei polimorfismi a singolo nucleotide presenti sulla regione di controllo e nella regione codificante, è stata gradualmente sostituita dall'analisi dell'intero genoma mitocondriale [123], che in ambito popolazionistico offre il vantaggio di una maggiore risoluzione, grazie ad un approfondito differenziamento di eventuali sotto-linee filogenetiche appartenenti allo stesso aplogruppo, come mostrato all'interno del nostro *set* di campioni attraverso l'incremento del numero di aplotipi prodotti e dalla maggiore quantità di aplotipi unici ricavati, insieme ad un drastico decremento degli aplotipi condivisi rispetto alla sola regione di controllo. In aggiunta, il maggiore potere discriminativo espresso in termini di *Haplotype diversity*, conferisce all'analisi dell'intero genoma la capacità di individualizzare due profili apparentemente accomunati dallo stesso aplogruppo ottenuto invece con l'analisi delle sole regioni ipervariabili HVS I e II. Questa caratteristica si rivela quindi molto utile ai fini dell'identificazione personale e negli studi di parentela in genetica forense, che necessita del più alto potere

discriminativo possibile nell'analisi della variabilità intra-individuale del DNA mitocondriale. In base a ciò, non sorprende il fatto che gli aplotipi ottenuti dall'intero mtDNA siano caratterizzati, rispetto alla regione di controllo, da una maggiore variabilità genetica espressa dall'alto valore di *Mean Pairwise Differences*, ed anche dal più basso valore ottenuto per la *Random Match Probability* definito come la probabilità di un *match* casuale tra due aplogruppi presenti all'interno di una popolazione e che può essere quindi inteso come la frequenza di un profilo genetico.

Un elemento importante questo che rivela l'importanza di includere all'interno di un *database* per il DNA mitocondriale, aplotipi originati dall'intera sequenza mtDNA per ottenere una maggiore informatività sul profilo analizzato.

Nessuna differenza nel rapporto tra transizioni e trasversioni di base è stato osservato a livello globale o nella sola regione di controllo. Questo risultato è concorde con la letteratura [124], così come la maggioranza osservata nella totalità dei campioni analizzati di transizioni rispetto al numero di trasversioni, in un rapporto stimato di 15:1 in linea con Rathburn *et.al* [125].

Per tutte le quattordici sessioni analitiche, è possibile notare che i valori espressi in percentuali di *usable reads* e quelli corrispondenti all'amplificazione monoclonale delle librerie genomiche rappresentati nella Tabella 6.3, sono tra loro inversamente proporzionali. Questa relazione è stata riportata recentemente in uno studio pubblicato da De Fanti *et al.* [126], in cui viene descritto che la percentuale di frammenti policlonali generati durante l'emulsione PCR per la presenza simultanea di due o più frammenti complessati alle biglie Ion PGM™ Hi-Q™ ISPs, è influenzata direttamente dal rapporto tra la quantità di biglie ISPs e le librerie da amplificare nella mix di PCR. Di conseguenza un maggiore controllo del rapporto tra questi due elementi, comporta un aumento del numero prodotto di *reads*, che si traduce in una migliore qualità del risultato.

Nonostante NGS sia un metodo tecnologicamente più raffinato rispetto a Sanger grazie al supporto di *plug-in* e *pipelines* esterne per la rilevazione delle varianti ed allineamento delle sequenze, l'analisi dei risultati richiede ancora il giudizio esperto per discriminare gli artefatti analitici sottoforma di inserzioni, delezioni o pseudo-eteroplasmie incluse erroneamente dal *plug-in variant-*

caller. Il 7% di errore commesso dal *plug-in* nella discriminazione di un artefatto analitico in posizione 249, 4824, 6641 e 8373, è imputabile all'utilizzo dei *primers* scelti nel pannello sperimentale Ampliseq Precision ID (Thermo Fisher Scientific) valutato nel presente studio, responsabili di introdurre pseudo-eteroplasmie in corrispondenza dei siti di annealing sulle estremità sovrapposte dei frammenti amplificati durante la fase di *emulsion* PCR.

Inoltre, la presenza di molteplici pseudo-eteroplasmie ad intensità variabile osservata per la regione 751-790 nell'allineamento della maggior parte dei campioni analizzati, ma non identificate dal *plug-in*, può essere attribuita alla co-amplificazione durante la PCR dei *Nuclear-encoded Mitochondrial Pseudogenes* (NUMTs), definiti come sequenze di DNA genomico non funzionale integrate all'interno del genoma nucleare, che possiedono una stretta omologia con i reali geni mitocondriali [127].

Il maggior numero di varianti di lunghezza osservate nella regione di controllo è attribuibile alla presenza di ampie sequenze *C-stretches* e dinucleotidiche nei tratti compresi tra le posizioni: 16184-16193, 303-309, 456-463, 515-523 e 568-573 che, come espresso da Forster *et al.* [64], per un valore superiore alle otto ripetizioni sono maggiormente prони a mutare il numero di basi. Complessivamente la quasi assoluta frequenza osservata nel campione per le varianti di lunghezza risulta in linea con la letteratura [62], che per la loro ricorrente presenza, vengono abitualmente non riportate nella ricostruzione filogenetica e nelle analisi forensi di *routine*.

Ad oggi la rilevazione delle eteroplasmie mediante Sanger è basata su una seconda analisi di conferma. Questo approccio però non è altrettanto applicabile al sequenziamento NGS per l'elevata variabilità nella produzione delle *reads* che richiede, invece, una soluzione basata su sistemi bioinformatici.

La corretta interpretazione e discriminazione delle misture eteroplasmiche rappresenta un importante punto critico per il genetista forense poiché la tecnologia NGS, caratterizzata da una maggiore sensibilità rispetto al sequenziamento Sanger, consente di rilevare segnali analitici

inferiori anche al 2% del *coverage*, che nel caso delle eteroplasmie puntiformi (PHP), potrebbero condurre ad una loro errata interpretazione.

Per questa ragione una posizione eteroplasmica, individuata dal *variant caller*, viene valutata ulteriormente attraverso l'adozione di un valore soglia corrispondente al rapporto tra la variante maggiore e minore rispetto al *coverage*, che viene definito *Minor Allele Frequency* (MAF). La scelta di un'adeguata soglia di discriminazione è importante poiché la condivisione di PHP tra due individui, fornisce al genetista forense un ulteriore supporto per una non esclusione all'interno di una analisi di parentela oppure in un quesito giudiziario.

E' stato infatti osservato che un MAF eccessivamente basso [128] offre risultati interpretativi controversi [129] perché risulterebbe inadeguato nel caso in cui le PHP presenti all'interno di un profilo, siano in realtà il prodotto di una mescolanza tra aplotipi originati da una contaminazione accidentale [130]. Nel nostro studio un MAF più elevato avrebbe ridotto ulteriormente il rumore di fondo analitico, ma avrebbe del pari causato la perdita di importanti informazioni per quelle eteroplasmie individuate con NGS e successivamente confermate in Sanger, le cui varianti minori sono presenti per il 15-16%.

Potendo escludere con certezza contaminazioni per i campioni Mg1_F09 ed Mg2_F09, è ipotizzabile che l'errata valutazione di due su 23 PHP osservate, possa essere invece causata dal basso *coverage* per le posizioni analizzate, rispettivamente di 74 e 66, perché, come riportato in precedenza, la qualità analitica è dipendente dalla quantità di reads prodotte.

Complessivamente, la percentuale di eteroplasmie puntiformi discriminata dal 15% di MAF nel nostro gruppo di individui risulta comunque sovrapponibile con quelle riportate in letteratura per Sanger ed NGS, per valori compresi tra il 10-20% [131].

Le discordanze riportate nel confronto col sequenziamento di Sanger rispettivamente in 78 campioni per la posizione 309.1C, in 2 campioni per le posizioni 563 e 16192 e in 4 campioni che differiscono per l'inserzione di una citosina in posizione 16193, si possono spiegare con la loro localizzazione all'interno delle regioni HVS I, II e III a livello dei tratti ripetuti di citosina "C-

stretches”, che per numero di nucleotidi ripetuti e complessità delle varianti di lunghezza, esercitano un impatto negativo sull’allineamento con la sequenza di riferimento di Cambridge, introducendo *bias* e difficoltà interpretative.

Anche se il metodo di Sanger non è esente dalla problematica delle *C-stretches*, le differenze osservate per Ion PGM System™ nel riportare il numero di nucleotidi inseriti e l’impossibilità di stabilirne con precisione la quantità riscontrate nella regione codificante per la posizione 8278 e nella regione di controllo per la posizione 563, potrebbero condurre all’ipotesi che il metodo analitico utilizzato dallo strumento NGS per il sequenziamento, basato sulla variazione del PH e della sua conversione in segnale digitale in proporzione al numero di nucleotidi incorporati, sia meno preciso all’aumentare del numero delle basi ripetute rispetto al metodo *gold-standard*. Ciò è mostrato nel confronto visivo per le due metodiche, riportato in Figura A.12 e A.13. Tuttavia, è necessario specificare che le posizioni 309.1C, 16192.1C e 16193.1C essendo caratterizzate da un’ampia variabilità e dalla quasi assoluta presenza all’interno delle differenti popolazioni, vengono escluse sia nella ricostruzione filogenetica, sia nella comparazione forense e nei database.

Per quanto riguarda invece le discordanze osservate con Ion PGM System™ rispettivamente per undici e cinque posizioni ritenute eteroplasmiche nei campioni Mg2_C04 e Mg2_C06 e la posizione 16274A, è possibile ipotizzare eventi di contaminazione avvenuti probabilmente durante la preparazione delle librerie genomiche, poiché sarebbe alquanto improbabile la presenza simultanea di più di due eteroplasmie puntiformi all’interno di uno stesso aplotipo.

Capitolo 8: Conclusioni

La capacità di sequenziare ampie regioni genomiche in tempi ridotti da parte della tecnologia *Next Generation Sequencing* è stata applicata con entusiasmo negli studi di *Genome Wide Analysis*, nelle diagnosi molecolari di malattie ereditarie, nella drug discovery, oppure in medicina legale nelle diagnosi delle cause di morte.

L'elevata sensibilità dello strumento, consente inoltre di superare i limiti imposti dalla degradazione del DNA, una caratteristica sfruttata in archeogenetica ed antropologia molecolare in cui molto spesso è necessario estrarre materiale genetico da reperti ossei. Questo aspetto in particolare, assieme agli stessi marcatori utilizzati in questi due ambiti scientifici, ha molto in comune con la genetica forense. Sebbene di recente siano già stati sviluppati pannelli NGS specifici per l'identificazione individuale tramite marcatori autosomici, cromosoma X ed Y (*HID-Ion AmpliSeq™ Identity panel*; ThermoFisher Scientific) (*ForenSeq™ DNA Signature Prep kit*; Illumina), nessun pannello è attualmente commercializzato per il DNA mitocondriale. Questo perché l'intera sequenza della molecola presenta alcune peculiarità come le regioni altamente ripetute, eteroplasmie, NUMTs, che possono complicare l'interpretazione del dato. In particolare, ancora poco si conosce riguardo al sequenziamento NGS basato su un sistema a semiconduttori, specialmente nei confronti di una molecola complessa come il mitogenoma che non è esente da problematiche interpretative nemmeno attraverso Sanger.

Nel seguente studio è stato condotto il confronto tra i due metodi di sequenziamento, valutando la presenza di eventuali differenze nell'analisi dell'intero genoma mitocondriale. Complessivamente entrambi i metodi non differiscono nell'analisi degli SNPs, sebbene la qualità del dato per NGS sia strettamente dipendente dal *coverage*. Le problematiche emerse riguardano principalmente la difficoltà nel valutare l'esatto numero di basi inserite all'interno delle varianti di lunghezza, ed in particolare, la corretta discriminazione delle eteroplasmie puntiformi (PHP). In questo caso

specifico, più che dal metodo di sequenziamento adottato, la causa è dipendente dalla grande sensibilità analitica dello strumento e dal differente rapporto tra le componenti *wild-type* e mutate, ad opera dei *bottleneck* genetici all'interno dell'ovocita. Tuttavia la corretta interpretazione delle PHP non è importante solamente dal punto di vista genetico-forense, dove la sua presenza rafforza l'associazione statistica di una condivisione del profilo, ma anche sotto l'aspetto clinico, la cui diagnosi è associata in diverse patologie mitocondriali. E' indubbio che siano necessari ulteriori studi in grado di fare luce sugli artefatti analitici prodotti dal sequenziatore NGS per ottenere una standardizzazione del metodo basato su criteri interpretativi che tengano in considerazione gli aspetti tecnici e biologici. Nel primo caso questi aspetti dovranno interessare gli algoritmi utilizzati per il riconoscimento delle varianti e dell'allineamento oltre ai kit per il sequenziamento NGS e cosa più importante: stabilire un'adeguata soglia di rilevazione per discriminare correttamente gli artefatti da eventuali eteroplasmie (MAF). Nel secondo invece, è necessario prendere in considerazione la variabilità delle eteroplasmie puntiformi in base al tessuto prelevato, al tasso evolutivo della posizione esaminata e più specificatamente, in base alla presenza degli *hotspot* mutazionali sul genoma mitocondriale.

Partendo dal presupposto che nessun metodo è esente da limitazioni, la capacità offerta dal *Next Generation Sequencing* di analizzare un elevato volume di dati in tempistiche ridotte, rappresenta un'importante rivoluzione dal punto di vista scientifico. Non ci sono dubbi quindi che il sequenziamento di Sanger cederà definitivamente il posto a queste tecnologie anche nelle analisi di *routine*, attraverso investimenti mirati all'ottimizzazione del dato analitico ed una disponibilità sempre crescente di pannelli specifici per i diversi ambiti di ricerca dell'area bio-medica ed antropologica.

Appendice

Mg1_A08 *H5ala* 146C;263G;315.1C;456T;523DEL;.524DEL.721C;750G;1438G;4336C;4769G;8860G;15326G;15833T;16304C

Mg1_B01 *H5a2* 263G;315.1C;456T;750G;1438G;3504G;4336C;4769G;5839T;8860G;14187T;15326G;16304C

Mg1_B06 *J1eDb* 73G;185A;228A;263G;295T;315.1C;462T;489C;750G;1438G;2706G;3010A;4216C;4769G;7028T;8860G;9120G;9632G;10398G;10685A;11251G;11719A;12083G;12612G;13708A;14766T;14798C;15326G;15452A;16069T;16126C

Mg1_B09 *H5al* 263G;315.1C;456T;523DEL;.524DEL.750G;1438G;4336C;4769G;8860G;15326G;15833T;16304C

Mg2_F10 *W6* 73G;189G;194T;195C;204C;207A;252C;263G;315.1C;709A;750G;1243C;1438G;2706G;3505G;4093G;4769G;5046A;5460A;7028T;8251A;8614C;8860G;8994A;11674T;11719A;11947G;12414C;12705T;14208C;14766T;15326T;15584C;16192T;16223T;16292T;16325C;16519C

Mg2_G01 *G1Ib2* 73G;295T;315.1C;462T;489C;750G;1438G;1733T;2706G;3010A;4216C;4769G;4991A;7028T;8269A;8860G;10398G;11251G;11919A;12612G;13708A;14766T;15326G;15452A;15978T;16069T;16126C;16145A;16222T;16261T

Mg2_G03 *U2eIh* 73G;217C;228A;263G;315.1C;340T;508G;750G;1438G;1811G;2706G;3720G;4769G;5390G;5426C;6045T;6152C;7028T;8860G;10876G;11467G;11719A;12308G;12372A;13020C;13734C;14766T;15326G;15907G;16051G;16129C;16183C;16193.1C;16362C;16519C

Mg2_G05 *H+146+195* 146C;195C;263G;315.1C;750G;1438G;4769G;6249A;8860G;9254G;15326G;16519C

Mg1_C09 *B4d1* 55C;56G;73G;263G;315.1C;750G;827G;1438G;2706G;4712T;4769G;7028T;8281DEL;.8282DEL;.8283DEL;.8284DEL;.8285DEL;.8286DEL;.8287DEL;.8288DEL;.8289DEL;8860G;11719A;11914A;13942G;14040A;14766T;15038G;15326G;15535T;15930A;16183C;16189Y;16217C;16519C

Mg1_C10 *X2+225* 73G;153G;195C;225A;263G;315.1C;750G;1438G;1719A;2706G;4769G;6221C;6371T;7028T;8860G;9815T;11719A;12705T;13966G;14470C;14544A;14766T;15326G;16129A;16189C;16223T;16278T;16344T;16519C

Mg1_D03 *H* 146C;263G;315.1C;750G;1438G;4502C;4769G;8860G;12662G;15326G;16172C;16390A;16519C

Mg1_E12 *H2ala* 73G;263G;315.1C;750G;951A;3335C;6173T;7076G;8860G;13095C;15326G;16354T

Mg1_F05 *H* 263G;315.1C;750G;1438G;4769G;6407C;8860G;8961G;15317A;15326G;16278T;16292T;16519C

Mg1_F07 *X2m2m* 16T;73G;153G;195C;225A;226C;263G;315.1C;750G;1438G;1719A;2706G;4769G;6221C;6371T;7028T;8860G;8966C;11719A;12454A;12705T;13966G;14470C;14766T;15326G;16183C;16189C;16193.1C;16223T;16248T;16278T;16519C

Mg1_F09 *H5ala* 263G;315.1C;456T;523DEL;.524DEL.721C;750G;1438G;4336C;4769G;8373A;8860G;15326G;15833T

Mg1_F11 *H6ala4* 239C;263G;315.1C;750G;1438G;3915A;4727G;4769G;8860G;9380A;10237C;11253C;15326G;16362C;16482G

Mg2_A04 *X2e2a* 73G;153G;195C;263G;267Y;315.1C;750G;1438G;1719A;2706G;3948G;4769G;6221C;6371T;7028T;8856A;8860G;11719A;12084T;12705T;13966G;14470C;14766T;15310C;15326G;16093C;16183C;16189C;16193.1C;16223T;16278T;16519C

Mg2_A05 *H3g3* 152C;263G;315.1C;750G;1438G;4769G;6776C;8860G;9196A;10754C;11581A;15326G;16519C

Mg1_2H_C04 *H35* 263G;315.1C;750G;1438G;3342T;4769G;8860G;8934T;15326G;16311C;16519C

Mg2_A07 *U5a2d* 73G;146C;263G;315.1C;489C;523DEL;.524DEL.750G;1438G;2706G;3197C;4769G;5263T;7028T;7843G;7853A;8860G;9204G;9477A;11467G;11719A;12308G;12372A;13617C;14287C;14766T;14793G;15301A;15326G;15559G;16192T;16256T;16270T;16278T;16526A

Mg2_A08 *T1al* 73G;152C;195C;263G;315.1C;709A;750G;1438G;1888A;2706G;3903T;4216C;4769G;4917G;7028T;8697A;8860G;9999C;10465C;11251G;11674Y;11719A;11947G;12633A;13368A;14766T;14905A;15326G;15452A;15607G;15928A;16126C;16163G;16186T;16189C;16294T;16519C

Mg2_B08 *HV13b* 263G;315.1C;750G;1438G;1654C;2706G;3290C;4769G;7028T;7697A;8860G;9377G;11152C;15326G;16184T;16291T;16357C;16519C

Mg2_B10 *X2Y* 73G;195C;257G;263G;315.1C;750G;1438G;1719A;2280T;2706G;4769G;5894G;6221C;6371T;7028T;8860G;11719A;12705T;13966G;14470C;14766T;15326G;16183C;16189C;16193.1C;16223T;16278T;16519C

Mg2_C02 *H6aIb2* 239C;263G;315.1C;750G;1438G;3915A;4727G;4769G;8860G;9098C;9254G;9380A;10589A;15326G;16311C;16362C;16482G

Mg2_C05 *Hij* 263G;315.1C;750G;1438G;1719A;3010A;4733C;4769G;8860G;8868C;15326G;16223T;16519C

Mg2_C07 *H5+16311* 73G;263G;315.1C;456T;750G;1438G;4769G;4961G;8860G;13759R;15326G;16213A;16300G;16304C;16311C

Mg2_C08 *H2b* 152C;263G;315.1C;408A;750G;4769G;8598C;8860G;13434G;15326G;16311C

Mg2_D01 *H2b* 152C;263G;315.1C;750G;4769G;8598C;8860G;10694T;15326G;16311C

Mg2_D02 *F2d* 73G;235G;249DEL;263G;297G;315.1C;574G;634C;709A;750G;1005C;1009T;1438G;1824C;2706G;3970T;4769G;6392C;7028T;7492T;7828G;8860G;9099T;10310A;10535C;10586A;11719A;12338C;13708A;13928C;14509G;14766T;15326G;16093C

Mg2_D04 *H1b* 263G;315.1C;750G;827G;1438G;3010A;4769G;8860G;12130C;15326G;16189C;16193.1C;16356C;16519C

Mg2_D11 *X2e2al* 73G;195C;263G;315.1C;750G;1438G;1719A;2706G;3948G;4769G;6221C;6371T;7028T;8860G;11719A;12084T;12705T;13327G;13966G;14470C;14766T;15310C;15326G;16182C;16183C;16189C;16193.1C;16223T;16278T;16519C

Mg1_E09 *U4bIa3a* 73G;146C;195C;263G;315.1C;499A;750G;1438G;1811G;2706G;3672G;4646C;4769G;5752DEL;.5999C;6047G;7028T;7265G;7705C;8860G;9989C;11332T;11339C;11467G;11674T;11719A;12308G;12372A;14620T;14766T;15326G;15693C;16172C;16356C;16519C

Mg1_G01 *H1ba* 263G;315.1C;750G;1438G;3010A;4769G;6023A;8860G;15326G;16270T;16519C

Mg2_E03 *H7b* 263G;315.1C;523DEL;.524DEL.750G;1438G;4769G;4793G;5348T;8860G;9661C;15326G;15607G;16519C

Mg2_E04 *H5al* 263G;315.1C;456T;523DEL;.524DEL.750G;1438G;4335Y;4336C;4769G;8860G;10634G;15326G;15833T;16304C;16311C

Mg2_E07 *H5* 263G;315.1C;456T;750G;1438G;4769G;8481T;8860G;9804A;15326G;16304C

Mg2_E10 *H1c.22* 263G;315.1C;477C;750G;1438G;3010A;4769G;8860G;10679G;15326G;16519C

Mg2_F07 *F4e* 73G;263G;315.1C;489C;573.1C;750G;1438G;2706G;3010A;3483A;4769G;4883T;5178A;7028T;8414T;8573A;8701G;8860G;9540C;9804A;10398G;110400T;10873C;11215T;11719A;12705T;14668T;14766T;14783C;14905A;15043A;15301A;15326G;16066G;16223T;16274A;16325C;16362C;16519C

Mg2_F08 *H1c* 263G;315.1C;477C;750G;1438G;3010A;4769G;8860G;15326G;16327T;16519C

Mg2_F11 *H2al* 263G;315.1C;750G;951A;8860G;15326G;16354T;16519C

Mg2_F12 *H2b* 152C;263G;315.1C;750G;4769G;8598C;8860G;10694T;15326G;16311C

Mg2_G04 *H6aIb2* 239C;263G;315.1C;750G;1438G;3915A;4727G;4769G;8860G;9254G;9380A;10589A;12756A;15326G;16362C;16482G

Mg1_A06 *B4d1a* 73G;152C;263G;315.1C;316A;750G;827G;1438G;2706G;2857C;4227G;4769G;7028T;8281del;8282del;8283del;8284del;8285del;8286del;8287del;8288del;8289del;8790A;8860G;10757G;11428A;11719A;11914A;12732C;13942G;14766T;15038G;15326G;15555T;15930A;16183C;16189C;16193.1C;16217C;16519C

Mg2_C12 *K2a5b* 73G;146C;152C;263G;280T;315.1C;324T;709A;750G;1438G;1811G;2706G;2831A;3480G;4561C;4769G;7028T;8860G;9055A;9698C;9716C;10550G;11151T;1299C;11467G;11719A;12308G;12372A;14167T;14766T;14798C;15326G;16224C;16311C;16519C

Mg1_B05 *H2* 263G;315.1C;750G;1438G;4769G;5483C;8255A;8860G;15326G;15727Y;16325C;16519C

Mg1_2H_D01 *H+152* 152C;263G;315.1C;750G;1438G;4769G;8860G;9230C;12290G;15326G

Mg2_G02 *H16* 263G;315.1C;750G;1438G;2706G;3507T;4769G;6755A;7028T;8860G;15326G;16172C;16311C

TABELLA A.1. Aplotipi ed aplogruppi ottenuti dall'analisi dell'intero genoma mitocondriale di 151 individui rappresentativi della popolazione Mongola

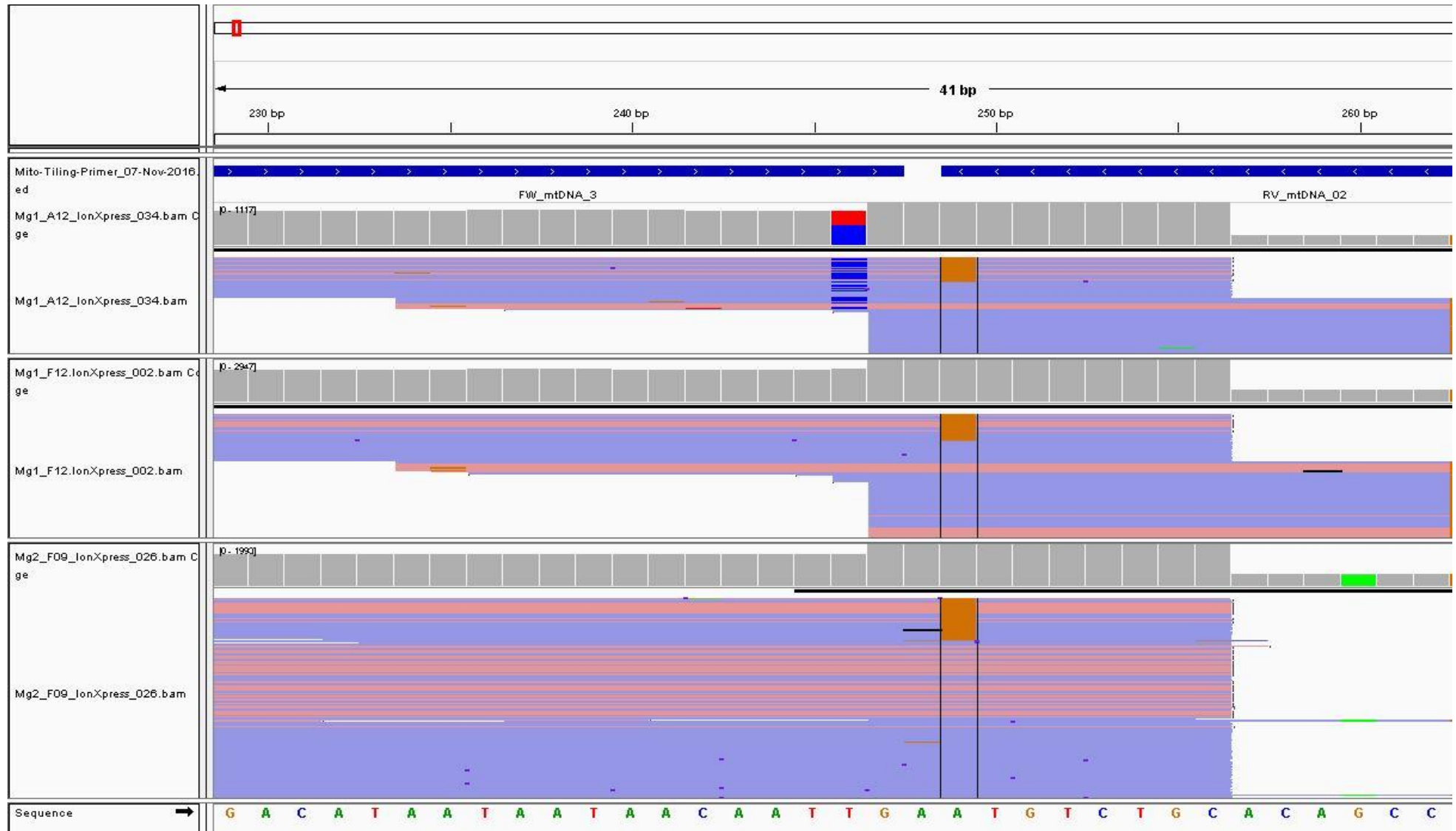


FIGURA A.1. Confronto simultaneo di un artefatto analitico in posizione 249 sotto forma di pseudo-eteroplasma



FIGURA A.2. Confronto simultaneo di un artefatto analitico erroneamente incluso dal *plug-in variant caller* in posizione 4824 sotto forma di pseudo-eteroplasma

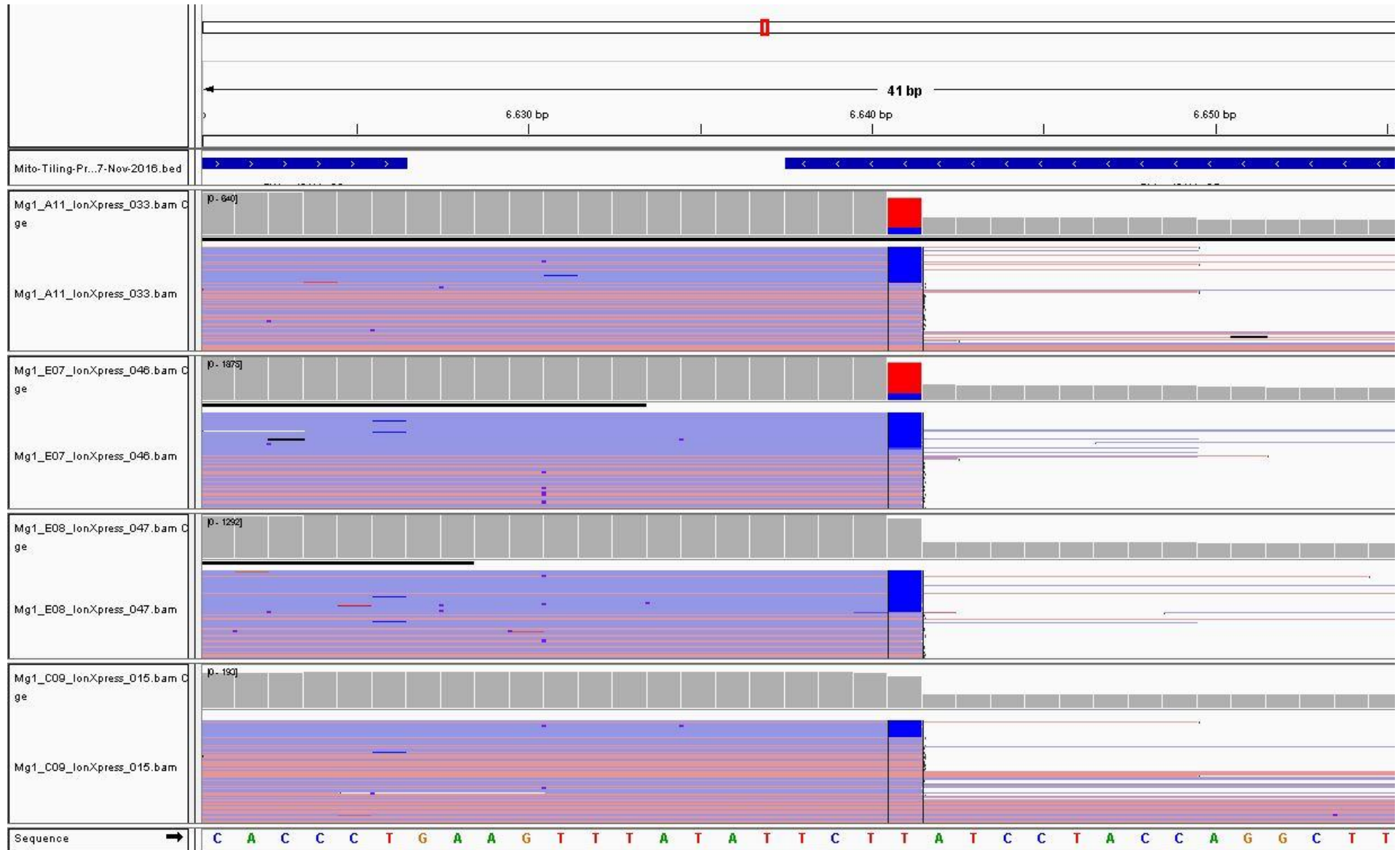


FIGURA A.3. Confronto simultaneo di un artefatto analitico erroneamente incluso dal *plug-in variant caller* in posizione 6641 sotto forma di pseudo-eteroplasma

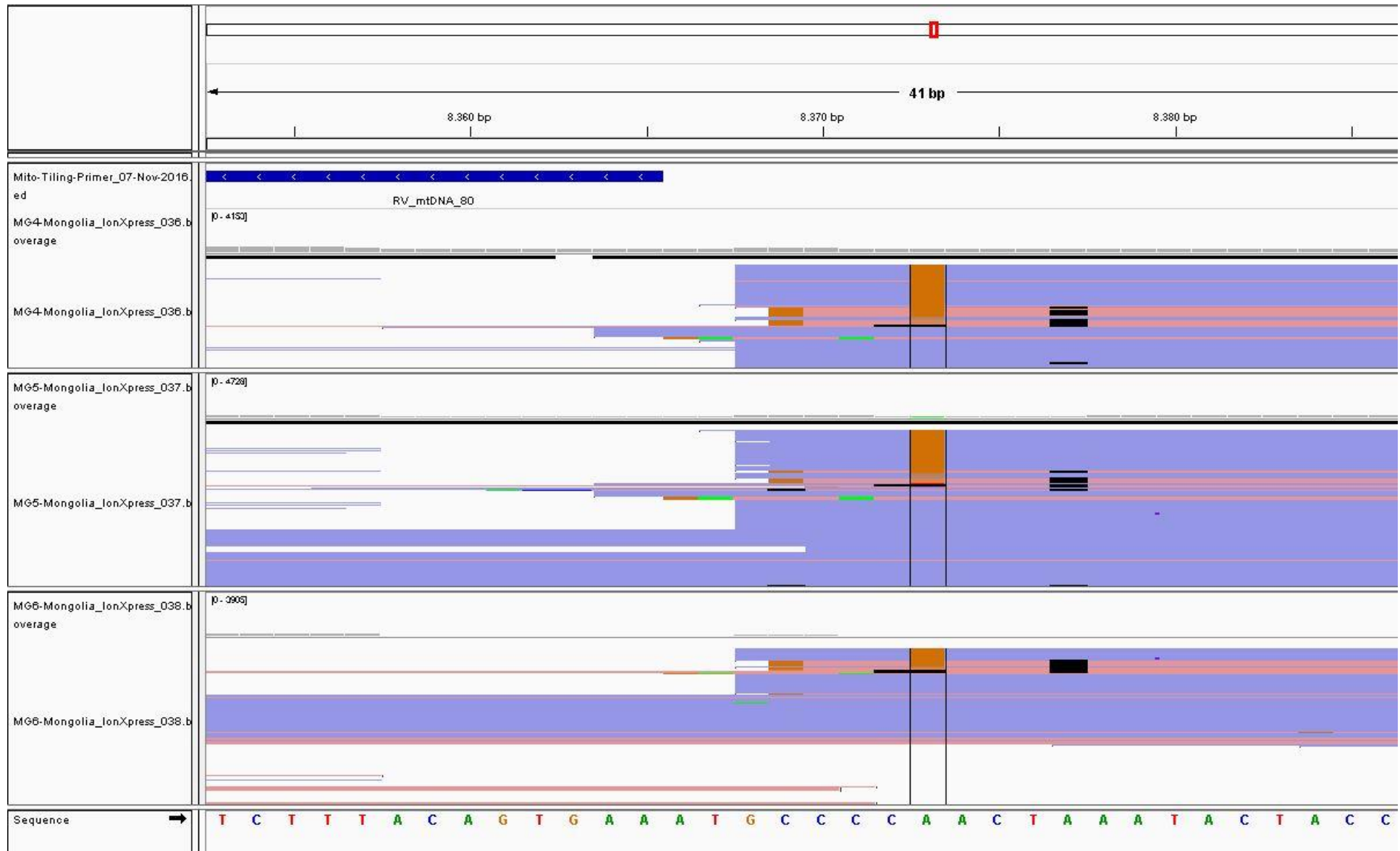


FIGURA A.4. Confronto simultaneo di un artefatto analitico erroneamente incluso dal *plug-in variant caller* in posizione 8373 sotto forma di pseudo-eteroplasma

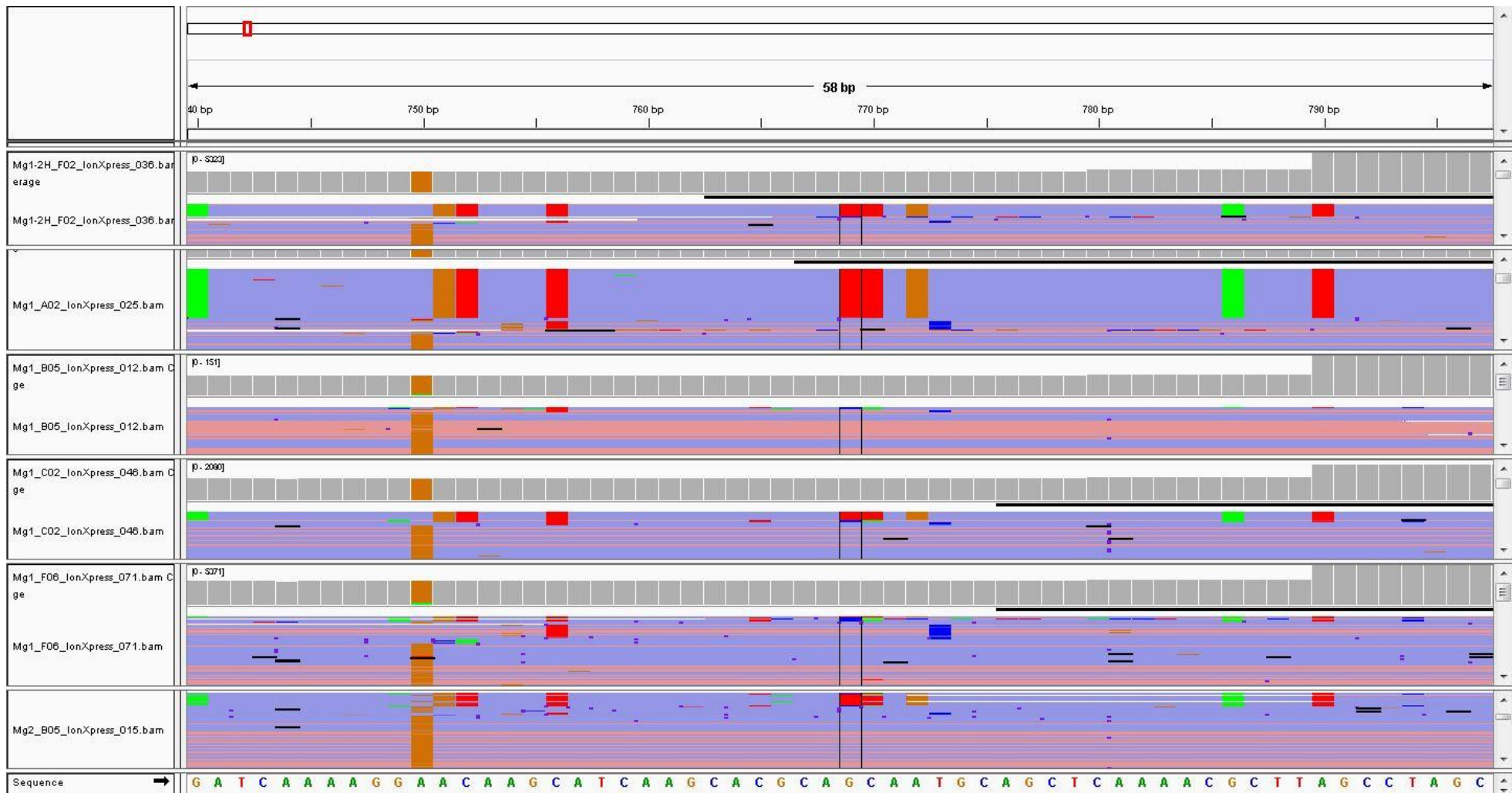


FIGURA A.5. Confronto simultaneo di artefatti analitici ad intensità variabili, non identificate dal *plug-in variant caller* presenti in tutti i 151 campioni analizzati

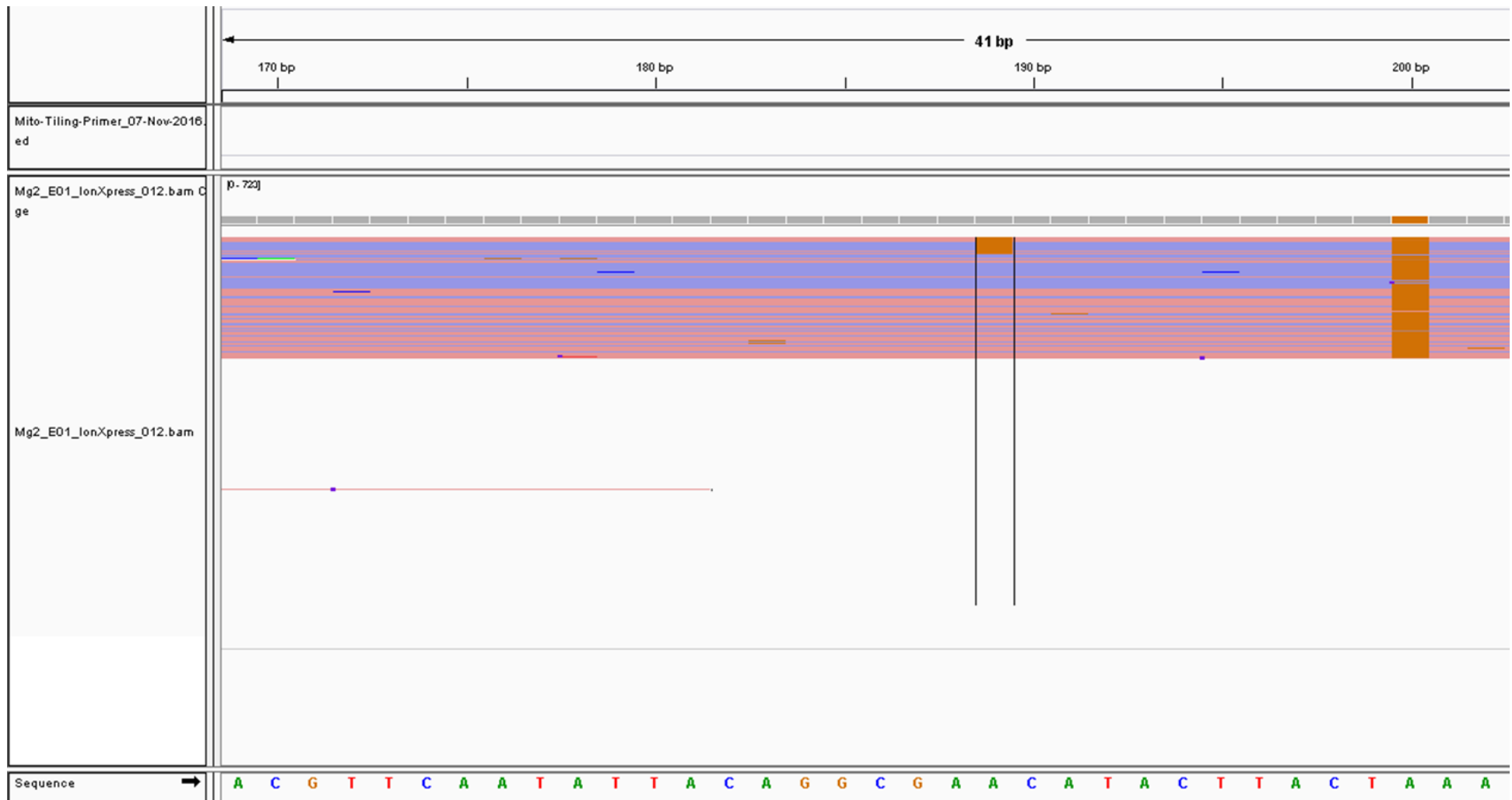


FIGURA A.6a. Allineamento per la posizione 189 del campione Mg2_E01

```

TTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTG:
TTTATCGCACCTACGTTCAATATTACAGGCGARCATACTTACTGAAGTGTGTTAATTAATTAATGCTTG:
TTTATCGCACCTACGTTCAATATTACAGGCGARCATACTTACTGAAGTGTGTTAATTAATTAATGCTTG:

```

```

| 160 | 170 | 180 | 190 | 200 | 210 | 220
TTTATCGCACCTACGTTCAATATTACAGGCGARCATACTTACTGAAGTGTGTTAATTAATTAATGCTTG:
+ .

```

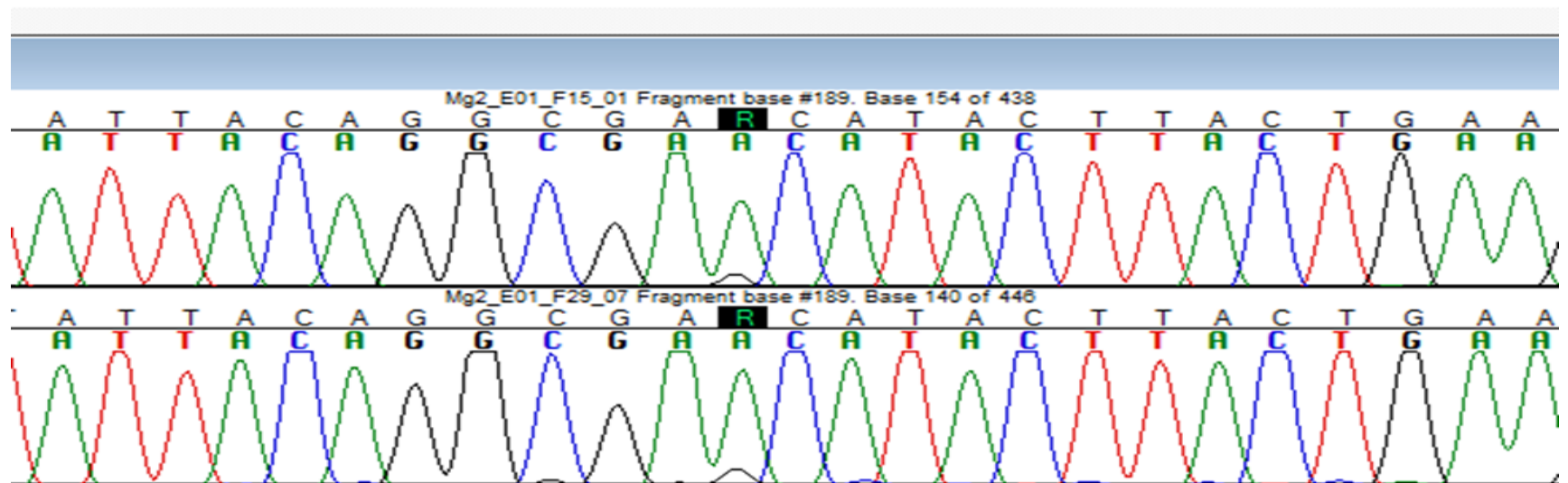


FIGURA A.6b. Conferma in Sanger del campione Mg2_E01 per la posizione 189

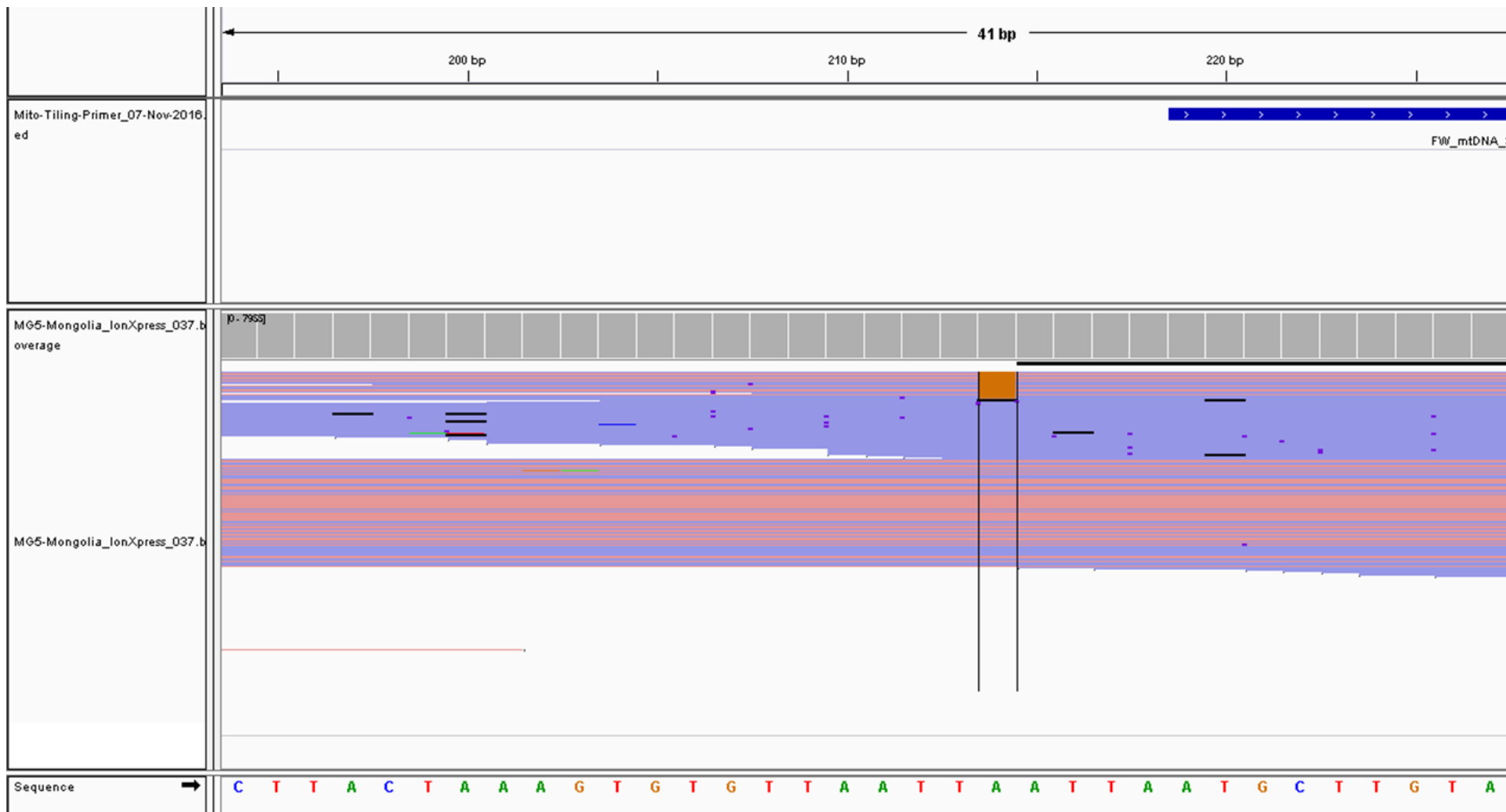


FIGURA A.7a. Allineamento per la posizione 214 del campione Mg5

GAACATACTTACTAAAGTGTGTTAATT**A**ATTAATGCTTGTAGGACATAATAAT
GAACATACTTACTAAAGTGTGTTAATT**A**ATTAATGCTTGTAGGACATAATAAT

190 200 210 220 230
GAACATACTTACTAAAGTGTGTTAATT**A**ATTAATGCTTGTAGGACATAATAAT
.....HV2.....

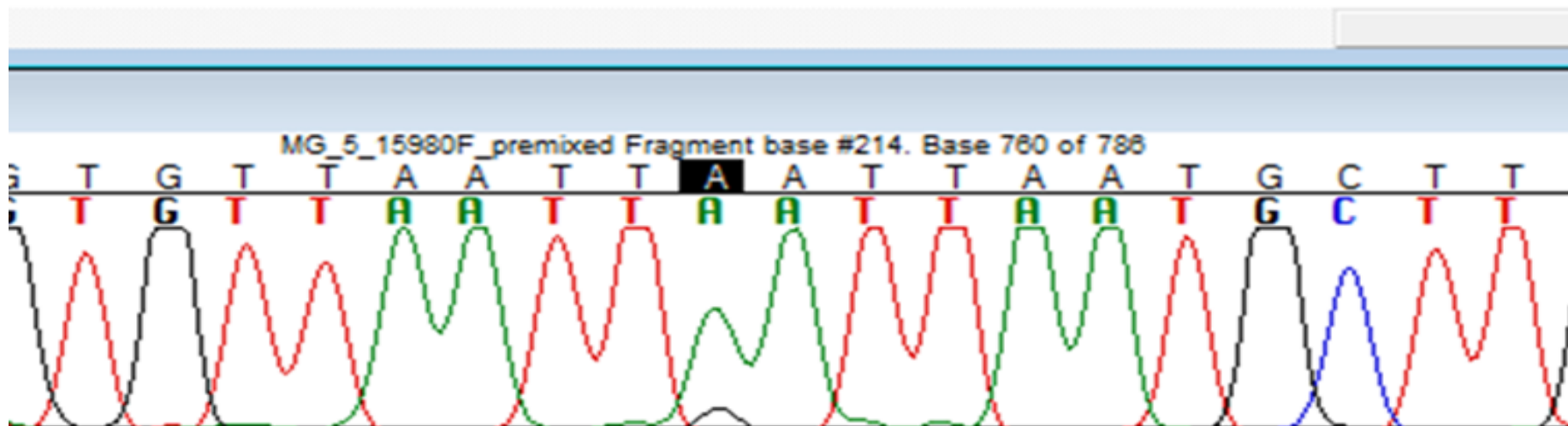


FIGURA A.7b. Conferma in Sanger del campioneMg5 per la posizione 214

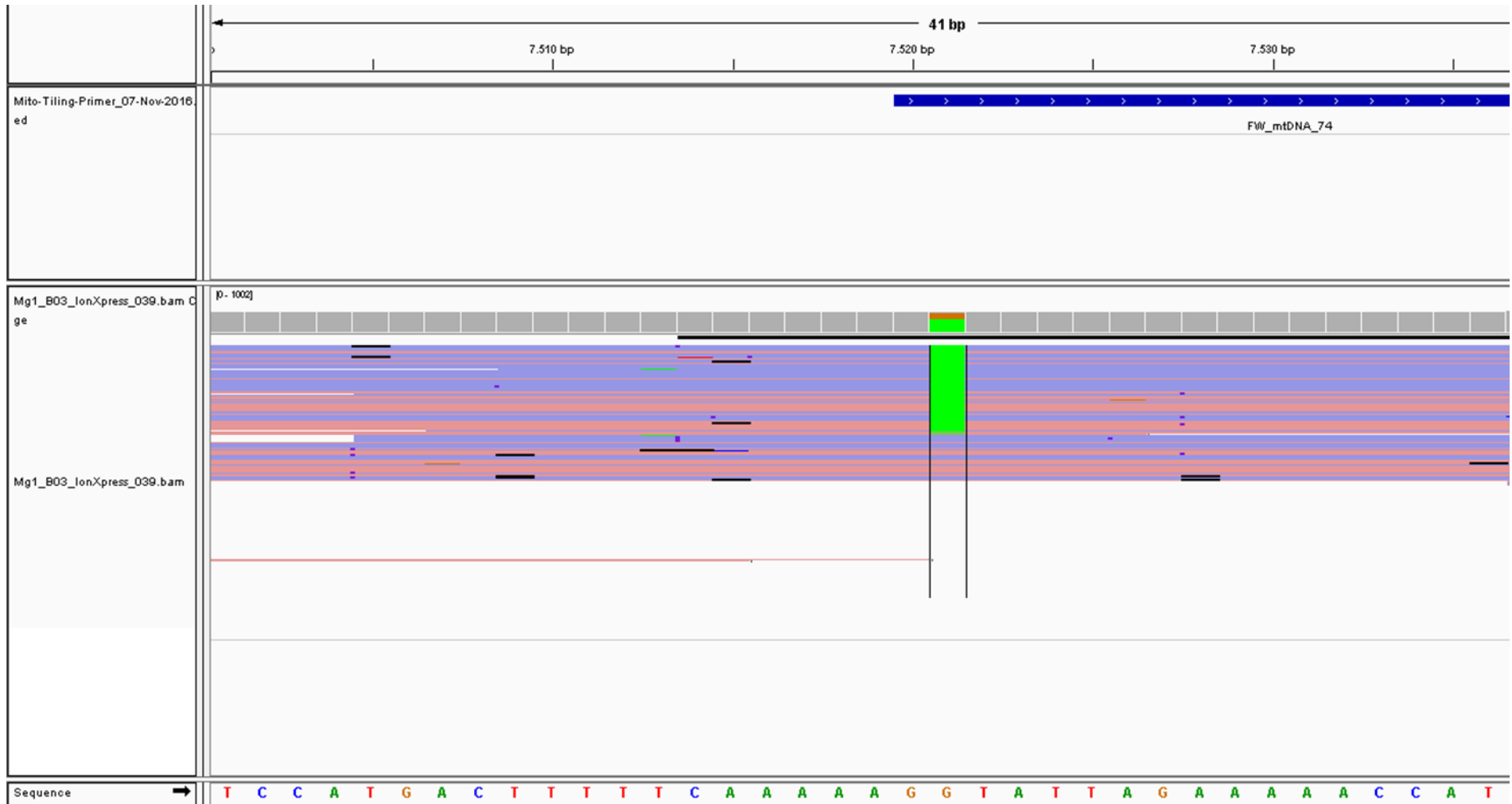


FIGURA A.8a. Allineamento per la posizione 7521 del campione Mg1_B03

```

TTTCAAGCCAACCCCATGGCCTCCATGACTTTTTTCAAAAAGG TATTAGAAAAACCATTTCATAACTTTGTCAAAGTTAAA
TTTCAAGCCAACCCCATGGCCTCCATGACTTT
TTTCAAGCCAACCCCATGGCCTCCATGACTTTTTTCAAAAAGR TATTAGAAAAACCATTTCATAACTTTGTCAAAGTTAAA
TTTCAAGCCAACCCCATGGCCTCCATGACTTTTTTCAAAAAGR TATTAGAAAAACCATTTCATAACTTTGTCAAAGTTAAA

```

```

7480 7490 7500 7510 7520 7530 7540 7550
TTTCAAGCCAACCCCATGGCCTCCATGACTTTTTTCAAAAAGR TATTAGAAAAACCATTTCATAACTTTGTCAAAGTTAAA
+

```

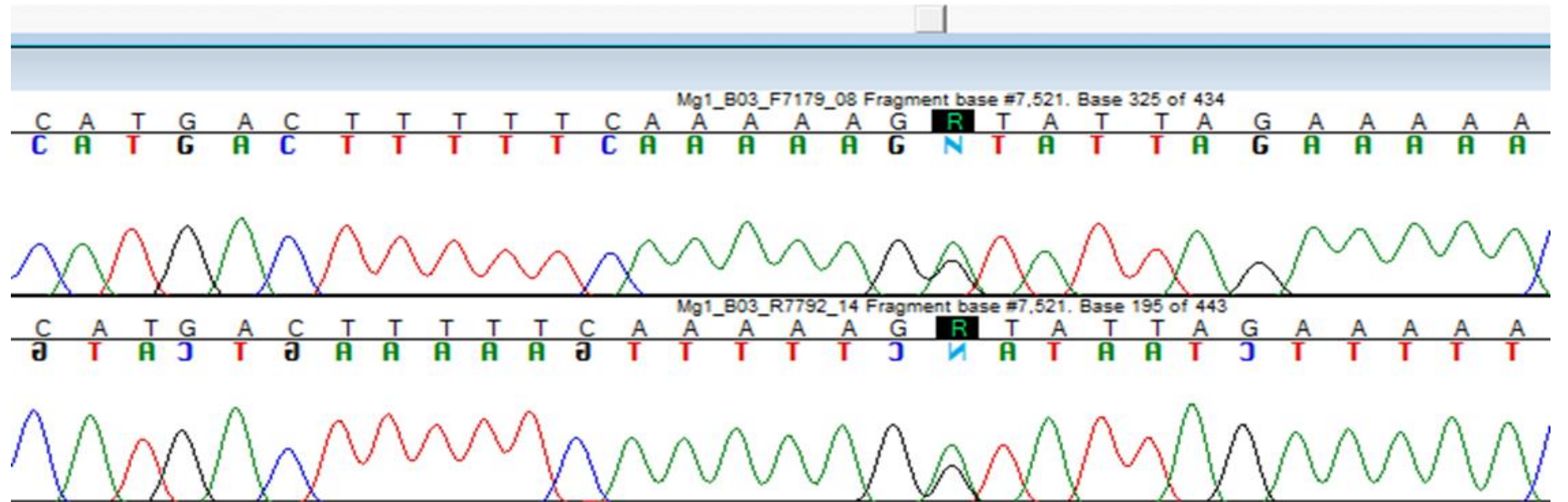


FIGURA A.8b. Conferma in Sanger del campione Mg1_B03 per la posizione 7521

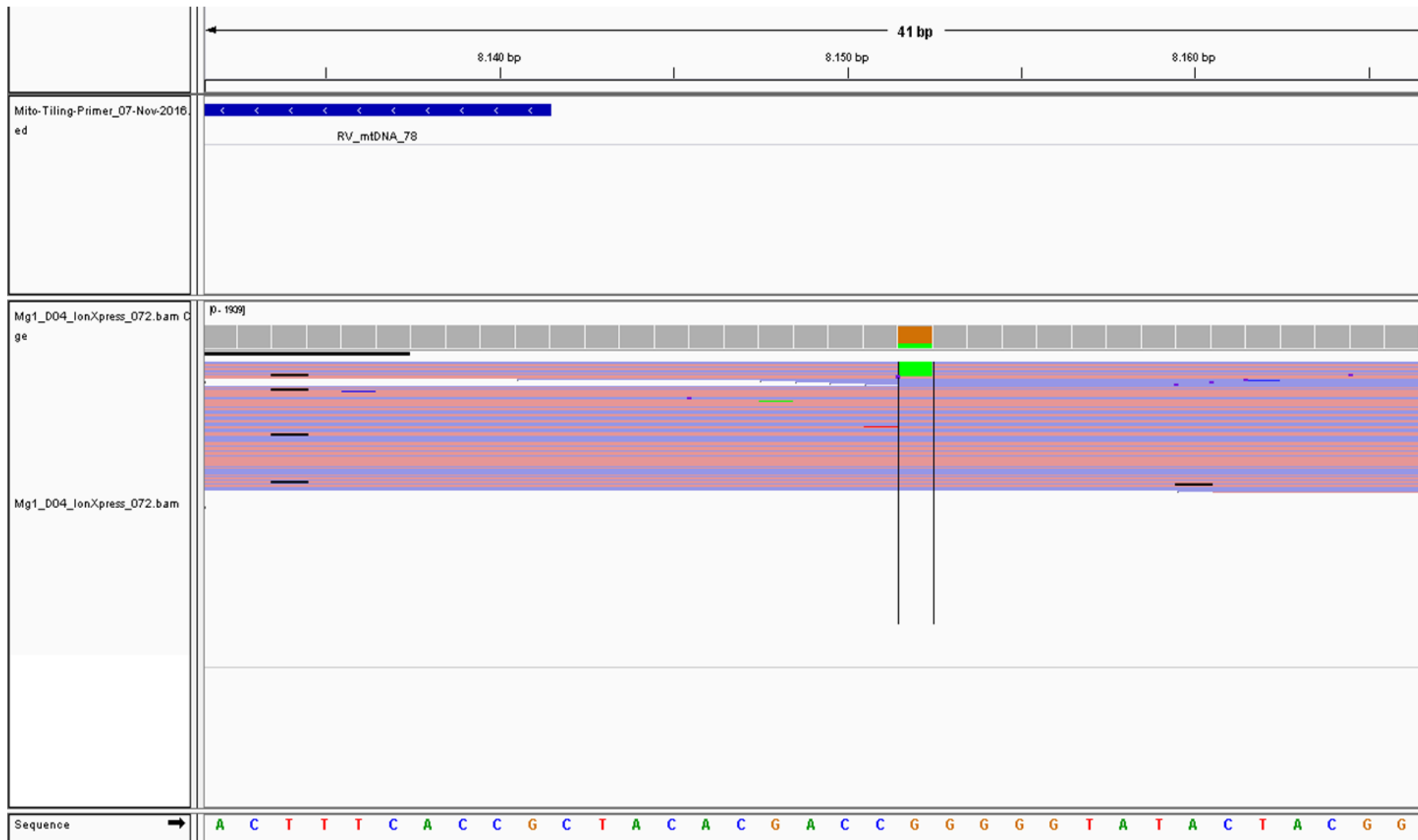


FIGURA A.9a. Allineamento per la posizione 8152 del campione Mg1_D04

```

BATGCAATTCCCGGACGTCTAAACCAAACCACTTTCACCGCTACACGACCGGGGGTATACTACGGTC
BATGCAATTCCCGGACGTCTAAACCAAACCACTTTCACCGCTACACGACCRGGGGTATACTACGGTC
BATGCAATTCCCGGACGTCTAAACCAAACCACTTTCACCGCTACACGACCRGGGGTATACTACGGTC
BATGCAATTCCCGGACGTCTAAACCAAACCACTTTCACCGCTACACGACCRGGGGTATACTACGGTC

```

```

100 | 8110 | 8120 | 8130 | 8140 | 8150 | 8160
BATGCAATTCCCGGACGTCTAAACCAAACCACTTTCACCGCTACACGACCRGGGGTATACTACGGTC
+
```

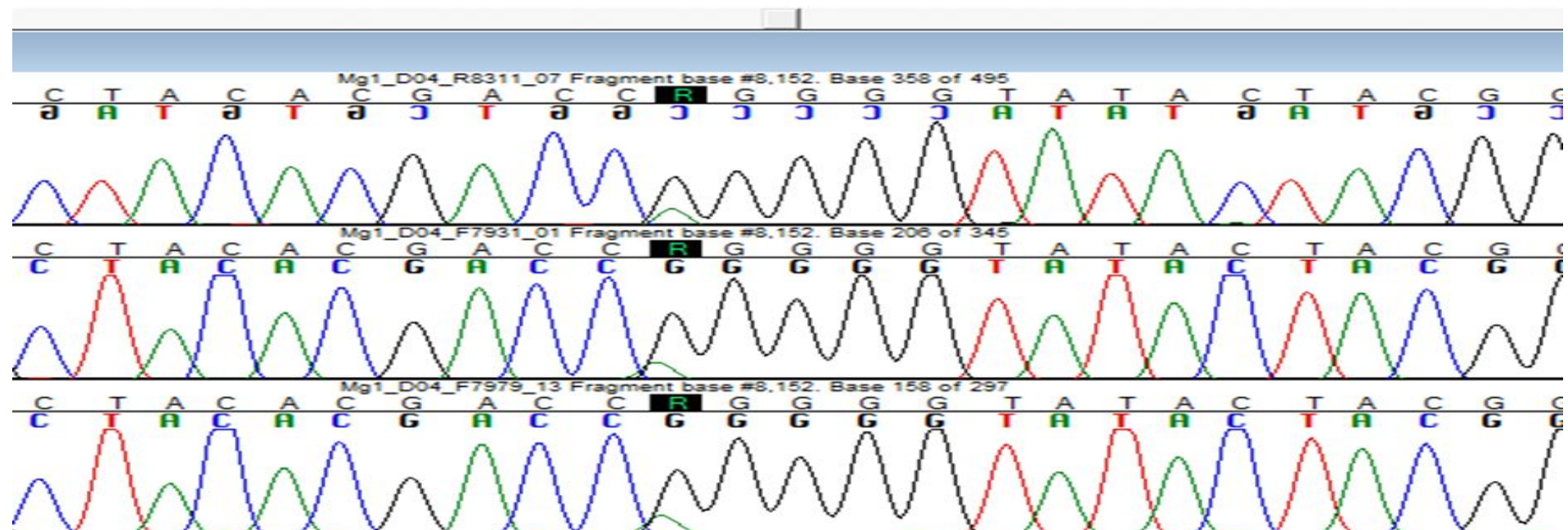


FIGURA A.9b. Conferma in Sanger del campione Mg1_D04 per la posizione 8152

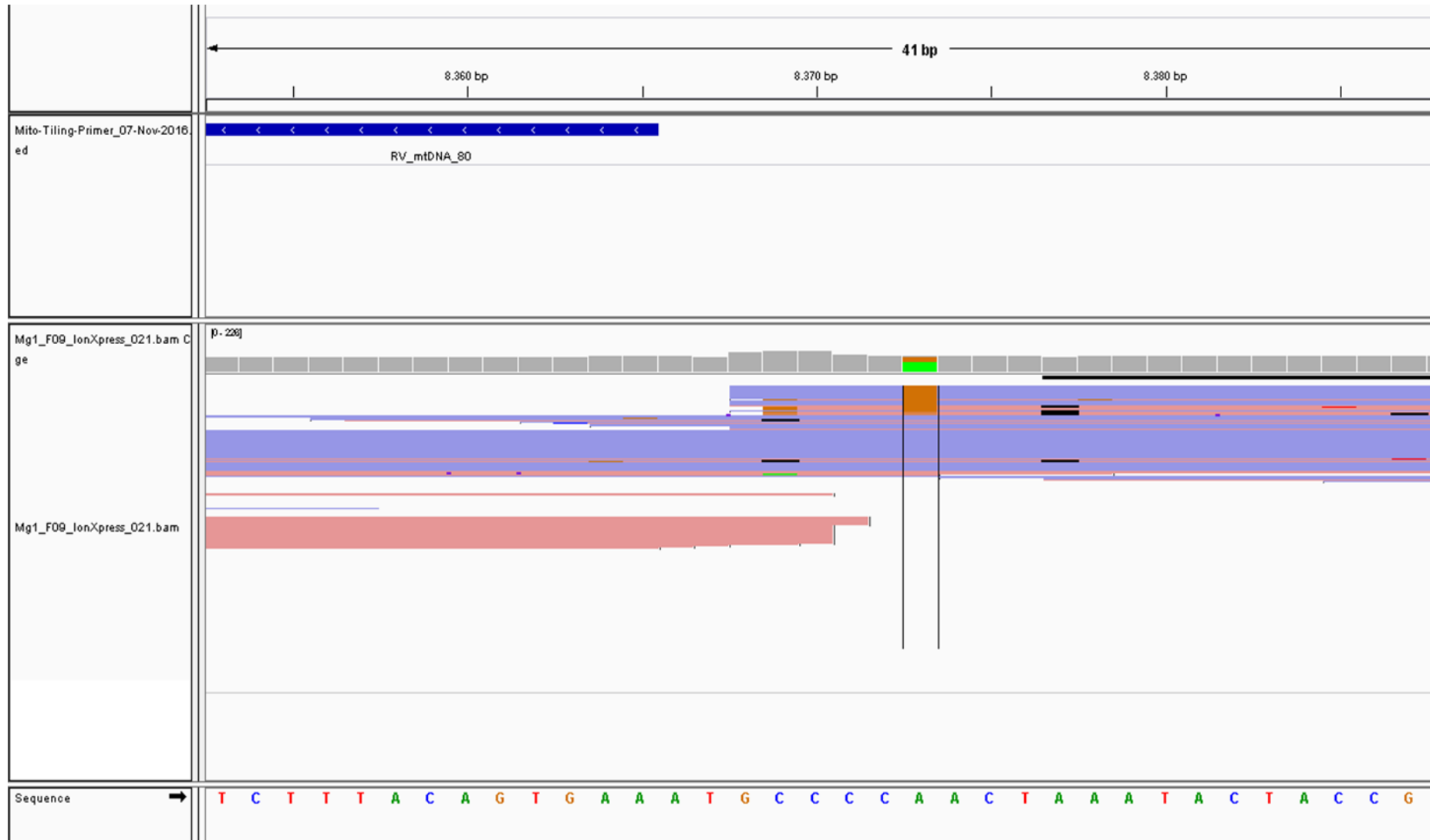


FIGURA A.10a. Allineamento per la posizione 8373 del campione Mg1_F09

```

GATTAAGAGAACCAACACCTCTTTACAGTGAAATGCCCCAACTAAATACTACCGTATGGCCCAC
GATTAAGAGAACCAACACCTCTTTACAGTGAAATGCCCCAACTAAATACTACCGTATGGCCCAC
GATTAAGAGAACCAACACCTCTTTACAGTGAAATGCCCCAACTAAATACTACCGTATGGCCCAC
GATTAAGAGAACCAACACCTCTTTACAGTGAAATGCCCCAACTAAATACTACCGTATGGCCCAC

```

```

0 | 8340 | 8350 | 8360 | 8370 | 8380 | 8390
GATTAAGAGAACCAACACCTCTTTACAGTGAAATGCCCCAACTAAATACTACCGTATGGCCCAC

```

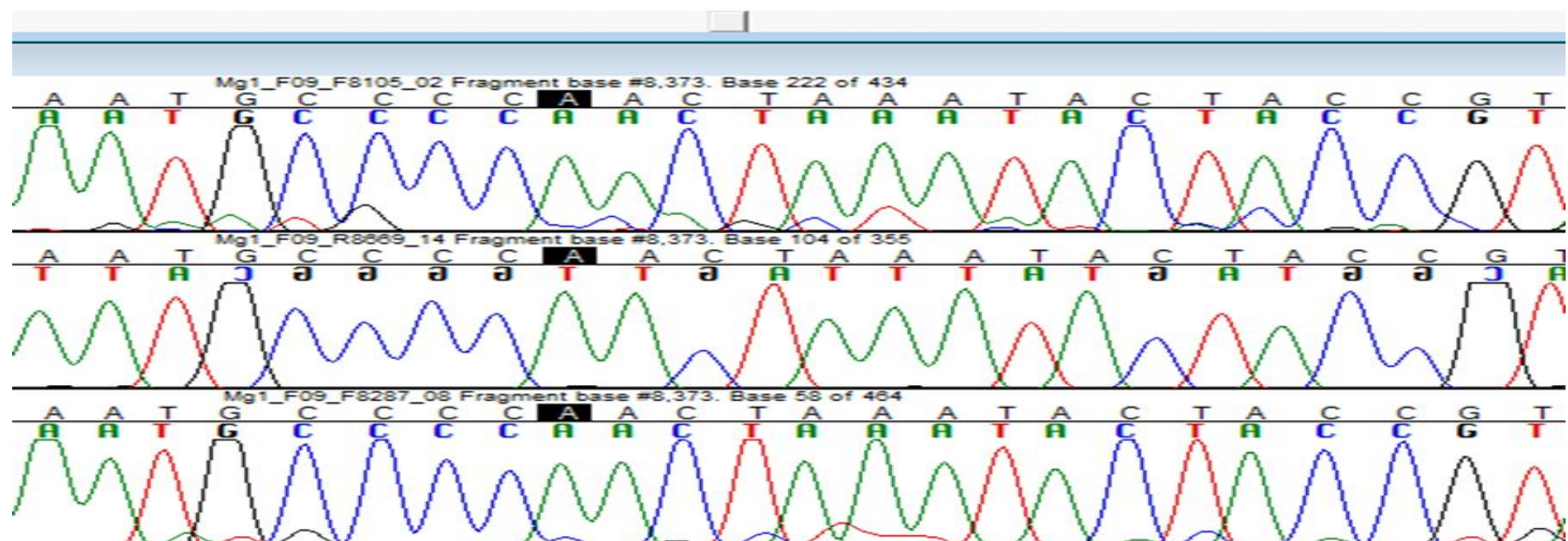


FIGURA A.10b. Sequenziamento di Sanger del campione Mg1_F09. Nessuna eteroplasmia puntiforme osservata per la posizione 8373

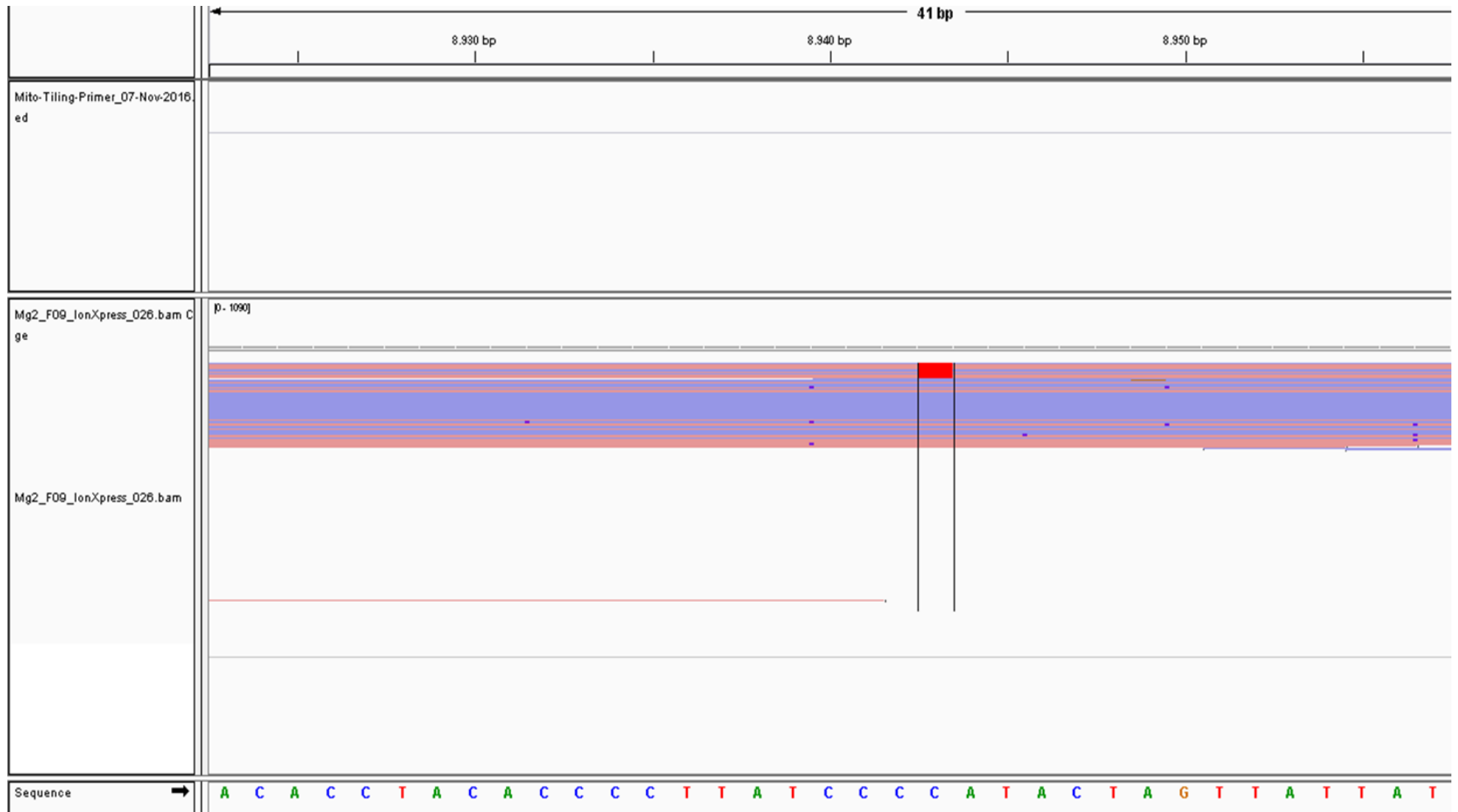


FIGURA A.11a. Allineamento per la posizione 8943 del campione Mg2_F09

```

AATGCCCTAGCCCACCTTCTTACCACAAGGCACACCTACACCCCTTATCCCCATACTAGTTATTATCGAAACCATCAGCCTACTCA
AATGCCCTAGCCCACCTTCTTACCACAAGGCACACCTACACCCCTTATCCCCATACTAGTTATTATCGAAACCATCAGCCTACTCA
TACACCCCTTATCCCCATACTAGTTATTATCGAAACCATCAGCCTACTCA
AGCCTACTCA

```

```

90 | 8900 | 8910 | 8920 | 8930 | 8940 | 8950 | 8960 | 8970
AATGCCCTAGCCCACCTTCTTACCACAAGGCACACCTACACCCCTTATCCCCATACTAGTTATTATCGAAACCATCAGCCTACTCA

```

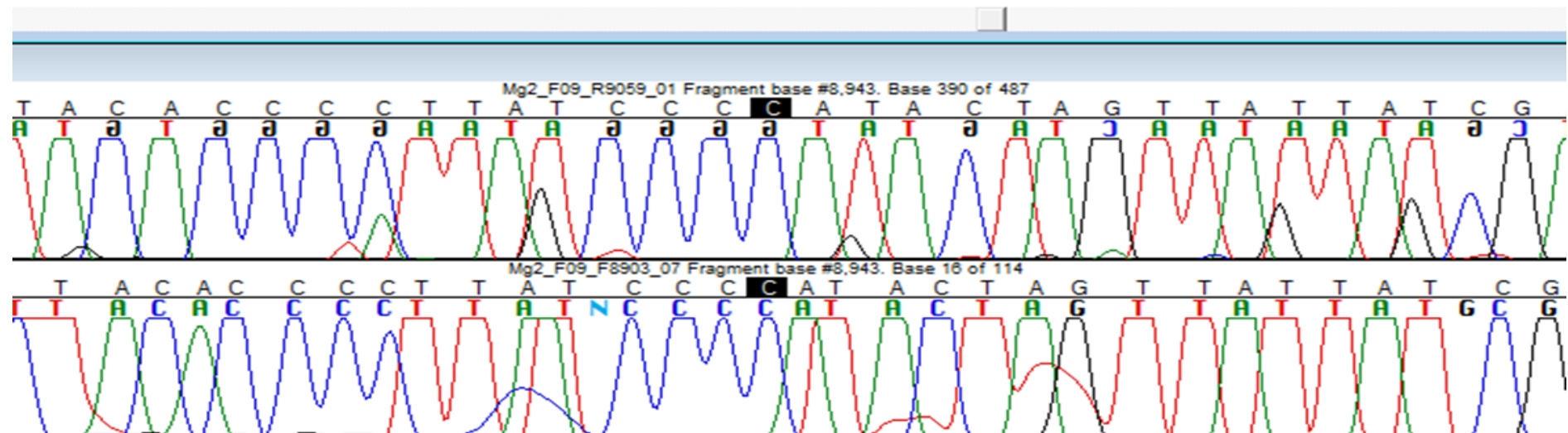
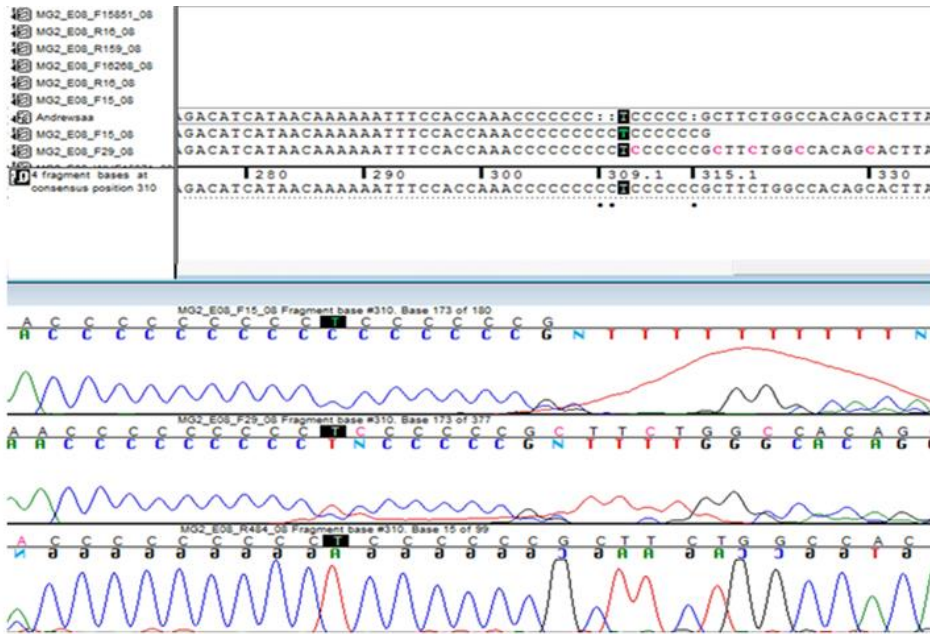


FIGURA A.11b. Sequenziamento di Sanger del campione Mg2_F09. Nessuna eteroplasmia puntiforme osservata per la posizione 8943

FIGURA A.12. Confronto tra gli allineamenti in Sanger (a) e metodo NGS (b) per il campione MG2_E08. Rispetto al sequenziamento NGS, Sanger si dimostra più preciso nell'analisi delle *C-stretches* per le posizioni 309.1C e 315.1C. Al contrario NGS non riesce a discriminare tra le due varianti di lunghezza, pertanto l'allineamento verrà interpretato, secondo le linee guida, riportando all'interno del profilo solamente la variante 315.1C.

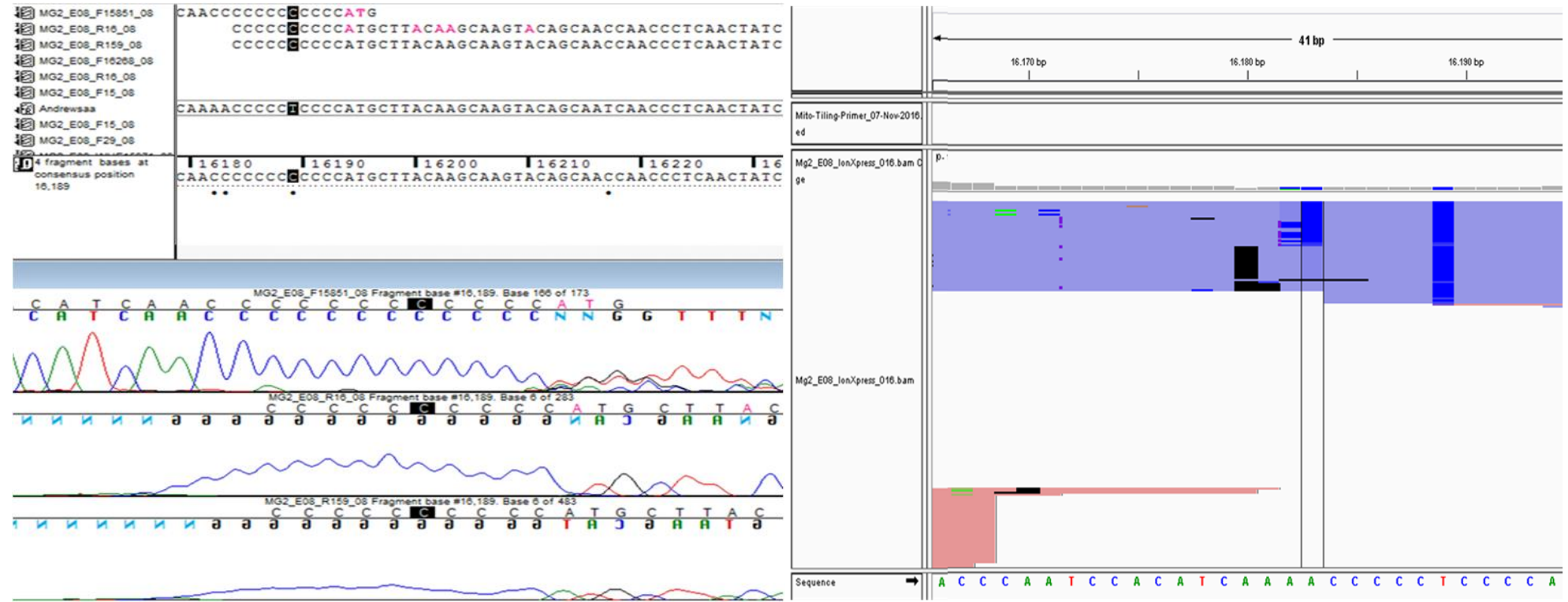


a)



b)

FIGURA A.13. Confronto tra gli allineamenti in Sanger (a) e metodo NGS (b) per il campione MG2_E08. Rispetto al sequenziamento NGS, Anche in questo caso Sanger si dimostra più preciso nell'individuare due trasversioni A/C a livello delle *C-stretch* nella regione 16180-16193, dove per NGS il segnale analitico risulta incompleto.



a)

b)

Bibliografia

- 1 Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature*. 1981 Apr 9;290(5806):457-65
- 2 Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature*. 2000 Dec 7;408(6813):708-13
- 3 Krings M, Stone A, Schlitz RW, Krainitzki H, Stoneking M, Paabo S. Neanderthal DNA sequences and the origin of modern humans. *Cell*. 1997;90(1):19-30
- 4 Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet*. 1996 Apr;12(4):417-20
- 5 Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*. 1988 Feb 25;331(6158):717-9
- 6 Richards M, Macaulay V. The mitochondrial gene tree comes of age. *Am J Hum Genet*. 2001 Jun;68(6):1315-20. PubMed PMID: 11349234
- 7 Lang BF, Gray MW, Burger G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet*. 1999;33:351-97. Review
- 8 . Chacinska, C.M. Koehler, D. Milenkovic, T. Lithgow, N. Pfanner, Importing mitochondrial proteins: machineries and mechanisms, *Cell* 138 (2009) 628-644
- 9 Rasmussen N. Mitochondrial structure and the practice of cell biology in the 1950s. *J Hist Biol*. 1995 Fall;28(3):381-429
- 10 Perkins G, Renken C, Martone ME, Young SJ, Ellisman M, Frey T. Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. *J Struct Biol*. 1997 Aug;119(3):260-72
- 11 Mannella CA, Pfeiffer DR, Bradshaw PC, Moraru II, Slepchenko B, Loew LM, Hsieh CE, Buttle K, Marko M. Topology of the mitochondrial inner membrane: dynamics and bioenergetic implications. *IUBMB Life*. 2001 Sep-Nov;52(3-5):93-100
- 12 Smith RA, Ord MJ. Mitochondrial form and function relationships in vivo: their potential in toxicology and pathology. *Int Rev Cytol*. 1983;83:63-134
- 13 Mannella CA. Introduction: our changing views of mitochondria. *J Bioenerg Biomembr*. 2000 Feb;32(1):1-4
- 14 Lea PJ, Hollenberg MJ. Mitochondrial structure revealed by high-resolution scanning electron microscopy. *Am J Anat*. 1989 Mar;184(3):245-57

-
- 15 Hackenbrock CR. Ultrastructural bases for metabolically linked mechanical activity in mitochondria. I. Reversible ultrastructural changes with change in metabolic steady state in isolated liver mitochondria. *J Cell Biol.* 1966 Aug;30(2):269-97
- 16 Frey TG, Mannella CA. The internal structure of mitochondria. *Trends Biochem Sci.* 2000 Jul;25(7):319-24. Review
- 17 Gonzalez-Freire M, de Cabo R, Bernier M, Sollott SJ, Fabbri E, Navas P, Ferrucci L. Reconsidering the Role of Mitochondria in Aging. *J Gerontol A Biol Sci Med Sci.* 2015 Nov;70(11):1334-42. doi: 10.1093/gerona/glv070. Review
- 18 Chistiakov DA, Sobenin IA, Revin VV, Orekhov AN, Bobryshev YV. Mitochondrial aging and age-related dysfunction of mitochondria. *Biomed Res Int.* 2014;2014:238463. doi: 10.1155/2014/238463. Review
- 19 Pollak, J.K., and Sutton, R. (1980) The differentiation of animal mitochondria during development. *Trends Biochem. Sci.*, 5:23-27.
- 20 Veltri KL, Espiritu M, Singh G. Distinct genomic copy number in mitochondria of different mammalian organs. *J Cell Physiol.* 1990 Apr;143(1):160-4
- 21 Lee HC, Wei YH. Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress. *Int J Biochem Cell Biol.* 2005 Apr;37(4):822-34. Review
- 22 Moyes, C. D., & Hood, D. A. (2003). Origins and consequences of mitochondrial variation in vertebrate muscle. *Annual Review of Physiology*,65, 177–201.
- 23 Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature.* 1981 Apr 9;290(5806):457-65
- 24 Szibor R, Michael M, Spitsyn VA, Plate I, Ginter EK, Krause D. Mitochondrial D-loop 3' (CA)_n repeat polymorphism: optimization of analysis and population data. *Electrophoresis.* 1997 Dec;18(15):2857-60
- 25 Richter C, Park JW, Ames BN. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci U S A.* 1988 Sep;85(17):6465-7
- 26 Brown TA, Tkachuk AN, Shtengel G, Kopek BG, Bogenhagen DF, Hess HF, Clayton DA. Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction. *Mol Cell Biol.* 2011 Dec;31(24):4994-5010. doi: 10.1128/MCB.05694-11. PubMed PMID: 22006021
- 27 Satoh M, Kuroiwa T. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res.* 1991 Sep;196(1):137-40
- 28 Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol.* 1988 Sep;136(3):507-13
- 29 Walberg MW, Clayton DA. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res.* 1981 Oct 24;9(20):5411-21

-
- 30 Shadel GS, Clayton DA. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem.* 1997;66:409-35
- 31 Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A.* 1980 Nov;77(11):6715-9
- 32 Ankel-Simons F, Cummins JM. Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proc Natl Acad Sci US A.* 1996 Nov 26;93(24):13859-63
- 33 Hoekstra RF. Evolutionary origin and consequences of uniparental mitochondrial inheritance. *Hum Reprod.* 2000 Jul;15 Suppl 2:102-11
- 34 F. Ankel-Simons, J.M. Cummins, Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution, *Proc. Natl.Acad. Sci. U. S. A.* 93 (1996) 13859–13863
- 35 U. Gyllensten, D. Wharton, A. Josefsson, A.C. Wilson, Paternal inheritance of mitochondrial DNA in mice, *Nature* 352 (1991) 255–257
- 36 H. Kaneda, J. Hayashi, S. Takahama, C. Taya, K.F. Lindahl, H. Yonekawa, Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis, *Proc. Natl. Acad. Sci. U. S. A.* 92 (1995) 4542–4546.
- 37 J.M. Cummins, T. Wakayama, R. Yanagimachi, Fate of microinjected sperm components in the mouse oocyte and embryo, *Zygote* 5 (1997) 301–308.
- 38 P. Sutovsky, R.D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly, G. Schatten, Ubiquitin tag for sperm mitochondria, *Nature* 402 (1999) 371–372
- 39 Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod.* 2000 Aug;63(2):582-90
- 40 Hurst LD, Hamilton WD (1992) Cytoplasmic fusion and the nature of sexes. *Proceedings of the Royal Society B: Biological Sciences*, 247, 189–194.
- 41 Allen JF (1996) Separate sexes and the mitochondrial theory of ageing. *Journal of Theoretical Biology*, 180, 135–140
- 42 Koehler CM, Lindberg GL, Brown DR, Beitz DC, Freeman AE, Mayfield JE, Myers AM. Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics.* 1991 Sep;129(1):247-55
- 43 Holt IJ, Miller DH, Harding AE (1989) Genetic heterogeneity and mitochondrial DNA heteroplasmy in Leber's hereditary optic neuropathy. *J Med Genet* 26:739–743
- 44 Vilkki J, Savontaus ML, Nikoskelainen EK (1990) Segregation of mitochondrial genomes in a heteroplasmic lineage with Leber hereditary optic neuroretinopathy. *Am J Hum Genet* 47:95–100

-
- 45 Cree, L. M. et al. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature Genet.* 40, 249–254 (2008)
- 46 Cao, L. et al. New evidence confirms that the mitochondrial bottleneck is generated without reduction of mitochondrial DNA content in early primordial germ cells of mice. *PLoS Genet.* 5, e1000756 (2009).
- 47 Wai, T., Teoli, D. & Shoubridge, E. A. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nature Genet.* 40, 1484–1488 (2008)
- 48 Sato, A. et al. Deletion-mutant mtDNA increases in somatic tissues but decreases in female germ cells with age. *Genetics* 177, 2031–2037 (2007)
- 49 Fan, W. et al. A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* 319, 958–962 (2008)
- 50 Stewart, J. B. et al. Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biol.* 6, e10 (2008)
- 51 Lee, H. S. et al. Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell Rep.* 1, 506–515 (2012).
- 52 Lee HS, Ma H, Juanes RC, Tachibana M, Sparman M, Woodward J, Ramsey C, Xu J, Kang EJ, Amato P, Mair G, Steinborn R, Mitalipov S. Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell Rep.* 2012 May 31;1(5):506-15
- 53 Melton T. Mitochondrial DNA Heteroplasmy. *Forensic Sci Rev.* 2004 Jan;16(1):1-20
- 54 Monnat RJ Jr, Loeb LA. Nucleotide sequence preservation of human mitochondrial DNA. *Proc Natl Acad Sci U S A.* 1985 May;82(9):2895-9
- 55 Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL. A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am J Hum Genet.* 2000 Aug;67(2):432-43
- 56 Bogenhagen DF. Repair of mtDNA in vertebrates. *Am J Hum Genet.* 1999 May;64(5):1276-81
- 57 Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. *Science.* 2006 Mar 24;311(5768):1727-30
- 58 Naue J, Hörer S, Sängler T, Strobl C, Hatzer-Grubwieser P, Parson W, Lutz-Bonengel S. Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion.* 2015 Jan;20:82-94
- 59 Chinnery PF, Zwijnenburg PJ, Walker M, Howell N, Taylor RW, Lightowlers RN, Bindoff L, Turnbull DM. Nonrandom tissue distribution of mutant mtDNA. *Am J Med Genet.* 1999 Aug 27;85(5):498-501

-
- 60 Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* 2011;12(6):R59
- 61 do Rosário Marinho AN, de Moraes MR, Santos S, Ribeiro-Dos-Santos A. Human aging and somatic point mutations in mtDNA: A comparative study of generational differences (grandparents and grandchildren). *Genet Mol Biol.* 2011 Jan;34(1):31-4
- 62 Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, Brandstätter A, Parson W, Parsons TJ. Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *J Mol Evol.* 2009 May;68(5):516-27
- 63 Bendall KE, Sykes BC. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet.* 1995 Aug;57(2):248-56
- 64 Forster L, Forster P, Gurney SM, Spencer M, Huang C, Röhl A, Brinkmann B. Evaluating length heteroplasmy in the human mitochondrial DNA control region. *Int J Legal Med.* 2010 Mar;124(2):133-42. doi: 10.1007/s00414-009-0385-0
- 65 Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet.* 1994 Oct;55(4):760-76
- 66 Walsh PS, Fildes NJ, Reynolds R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* 1996 Jul 15;24(14):2807-12
- 67 Meyer S, Weiss G, von Haeseler A. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics.* 1999 Jul;152(3):1103-10
- 68 Pääbo S. Mutational hot spots in the mitochondrial microcosm. *Am J Hum Genet.* 1996 Sep;59(3):493-6
- 69 Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet.* 1997 Apr;15(4):363-8
- 70 Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet.* 2000 Apr;66(4):1384-97
- 71 Brandstätter A, Parson W. Mitochondrial DNA heteroplasmy or artefacts—a matter of the amplification strategy? *Int J Legal Med.* 2003 Jun;117(3):180-4
- 72 Forster L, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B. Natural radioactivity and human mitochondrial DNA mutations. *Proc Natl Acad Sci U S A.* 2002 Oct 15;99(21):13950-4

-
- 73 Rose G, Romeo G, Dato S, Crocco P, Bruni AC, Hervonen A, Majamaa K, Sevini F, Franceschi C, Passarino G; GENetics of Healthy Ageing Project Consortium. Somatic point mutations in mtDNA control region are influenced by genetic background and associated with healthy aging: a GEHA study. *PLoS One*. 2010 Oct 14;5(10)
- 74 He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*. 2010 Mar 25;464(7288):610-4
- 75 Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*. 1993 Sep;53(3):563-90
- 76 Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, Savontaus ML, Huoponen K, Scozzari R, Bandelt HJ. mtDNA haplogroups and frequency patterns in Europe. *Am J Hum Genet*. 2000 Mar;66(3):1173-7
- 77 Brown WM. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci U S A*. 1980 Jun;77(6):3605-9
- 78 Excoffier L, Langaney A. Origin and differentiation of human mitochondrial DNA. *Am J Hum Genet*. 1989 Jan;44(1):73-85
- 79 Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature*. 1987 Jan 1-7;325(6099):31-6
- 80 Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, Wallace DC. Ethnic variation in Hpa I endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1981 Sep;78(9):5768-72
- 81 Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet*. 1994 Oct;55(4):760-76
- 82 Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol*. 1994 Feb;93(2):189-99
- 83 Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli EL, Silva NM, Kivisild T, Torroni A, Villems R. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet*. 2012 Apr 6;90(4):675-84
- 84 Cerezo M, Achilli A, Olivieri A, Perego UA, Gómez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, López-Soto M, Carracedo A, Capelli C, Torroni A, Salas A. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res*. 2012 May;22(5):821-6
- 85 Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet*. 2004 Nov;75(5):752-70. Erratum in: *Am J Hum Genet*. 2006 Jun;78(6):1097

-
- 86 Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards MB. The archaeogenetics of Europe. *Curr Biol.* 2010 Feb 23;20(4):R174-83
- 87 Chaubey G, Metspalu M, Kivisild T, Villems R. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays.* 2007 Jan;29(1):91-100
- 88 Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, Cox MP. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *J Hum Genet.* 2013 Mar;58(3):165-73
- 89 Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, van Oven M, Behar DM, Rootsi S, Hudjashov G, Mallick CB, Karmin M, Nelis M, Parik J, Reddy AG, Metspalu E, van Driem G, Xue Y, Tyler-Smith C, Thangaraj K, Singh L, Remm M, Richards MB, Lahr MM, Kayser M, Villems R, Kivisild T. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol.* 2011 Feb;28(2):1013-24
- 90 Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, Sanz P, Yang-Yao D, Manry J, Ciani G, Luiselli D, Quintana-Murci L, Comas D, Pettener D; Genographic Consortium.. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS One.* 2013 May 29;8(5):e65441
- 91 Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, Nordenfelt S, Valdiosera C, Richards SM, Rohrlach A, Romero MI, Espinoza IF, Cagigao ET, Jiménez LW, Makowski K, Reyna IS, Lory JM, Torrez JA, Rivera MA, Burger RL, Ceruti MC, Reinhard J, Wells RS, Politis G, Santoro CM, Standen VG, Smith C, Reich D, Ho SY, Cooper A, Haak W. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci Adv.* 2016 Apr 1;2(4):e1501385
- 92 Capelli C, Tschentscher F, Pascali VL. "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient DNA analysis. *Forensic Sci Int.* 2003 Jan 9;131(1):59-64
- 93 Salas A, Bandelt HJ, Macaulay V, Richards MB. Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci Int.* 2007 May 3;168(1):1-13
- 94 Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL. Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int.* 1999 Jul 12;103(1):23-35
- 95 Forster P. To err is human. *Ann Hum Genet.* 2003 Jan;67(Pt 1):2-4
- 96 Röck AW, Dür A, van Oven M, Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Sci Int Genet.* 2013 Dec;7(6):601-9
- 97 Parson W, Gusmão L, Hares DR, Irwin JA, Mayr WR, Morling N, Pokorak E, Prinz M, Salas A, Schneider PM, Parsons TJ; DNA Commission of the International Society for Forensic Genetics.. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci Int Genet.* 2014 Nov;13:134-42

-
- 98 van den Ouweland JM, Lemkes HH, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, van de Kamp JJ, Maassen JA. Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. *Nat Genet* 1992;1:368–371
- 99 Goto Y, Nonaka I, Horai S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* 1990;348:651–653
- 100 Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am J Hum Genet* 1990;46:428–433
- 101 Wallace DC, Zheng XX, Lott MT, Shoffner JM, Hodge JA, Kelley RI, Epstein CM, Hopkins LC. Familial mitochondrial encephalomyopathy (MERRF): genetic, pathophysiological, and biochemical characterization of a mitochondrial DNA disease. *Cell* 1988b;55:601–610
- 102 Johns DR, Neufeld MJ, Park RD. An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy. *Biochem Biophys Res Commun* 1992;187:1551–1557
- 103 Wong LJ. Pathogenic mitochondrial DNA mutations in protein-coding genes. *Muscle Nerve*. 2007 Sep;36(3):279–93
- 104 M.-E. Patti, S. Corvera, The role of mitochondria in the pathogenesis of type 2 diabetes, *Endocr. Rev.* 31 (2010) 364–395.
- 105 D.C. Wallace, Mitochondria and cancer, *Nat. Rev. Cancer* 12 (2012) 685–698
- 106 I. Piaceri, V. Rinnoci, S. Bagnoli, Y. Failli, S. Sorbi, Mitochondria and Alzheimer's disease, *J. Neurol. Sci.* 322 (2012) 31–34
- 107 Y. Mizuno, S. Ikebe, N. Hattori, Y. Nakagawahattori, H. Mochizuki, M. Tanaka, T. Ozawa, Role of mitochondria in the etiology and pathogenesis of parkinsons disease, *Biochim. Biophys. Acta Mol. Basis Dis.* 1271 (1995) 265–274
- 108 Jeppesen TD, Schwartz M, Frederiksen AL, Wibrand F, Olsen DB, Vissing J. Muscle phenotype and mutation load in 51 persons with the 3243A.G mitochondrial DNA mutation. *Arch Neurol* 2006;63:1701–1706.
- 109 Rossignol R, Malgat M, Mazat JP, Letellier T. Threshold effect and tissues specificity. Implication for mitochondrial cytopathies. *J Biol Chem.* 1999 Nov 19;274(47):33426–32
- 110 Wallace DC. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* 2005; 39: 359–407.
- 111 Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 2004; 303:223–226.
- 112 Pyle A, Foltynie T, Tiangyou W et al. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. *Ann Neurol* 2005; 57: 564–567

-
- 113 Burgstaller JP, Johnston IG, Poulton J. Mitochondrial DNA disease and developmental implications for reproductive strategies. *Mol Hum Reprod.* 2015 Jan;21(1):11-22
- 114 Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008 May;18(5):821-9
- 115 Howell N, McCullough DA, Kubacka I, Halvorson S, Mackey D. The sequence of human mtDNA: the question of errors versus polymorphisms. *Am J Hum Genet.* 1992 Jun;50(6):1333-40
- 116 Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform.* 2010 Sep;11(5):473-83
- 117 H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Briefings Bioinform.* 14 (2013) 178–192.
- 118 van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009 Feb;30(2):E386-94
- 119 Rosenblum, B.B., Oaks, F., Menchen, S. and Johnson, B. (1997) *Nucleic Acids Research* 25, 3925–3929
- 120 Excoffier, L. G. Laval, and S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50
- 121 Cheng B, Tang W, He L, Dong Y, Lu J, Lei Y, Yu H, Zhang J, Xiao C. Genetic imprint of the Mongol: signal from phylogeographic analysis of mitochondrial DNA. *J Hum Genet.* 2008;53(10):905-13
- 122 Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol Biol Evol.* 2004 Dec;21(12):2265-80
- 123 Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, Bobillo MC, Corach D, Woodward SR, Salas A, Achilli A, Torroni A, Bandelt HJ, Parson W. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* 2012 May;22(5):811-20
- 124 Kristina Strandberg AK, Salter LA. A comparison of methods for estimating the transition:transversion ratio from DNA sequences. *Mol Phylogenet Evol.* 2004 Aug;32(2):495-503
- 125 Rathbun MM, McElhoe JA, Parson W, Holland MM. Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data. *Forensic Sci Int Genet.* 2017
- 126 De Fanti S, Vianello D, Giuliani C, Quagliariello A, Cherubini A, Sevini F, Iaquilano N, Franceschi C, Sazzini M, Luiselli D. Massive parallel sequencing of human whole mitochondrial genomes with Ion Torrent technology: an optimized workflow for Anthropological and Population Genetics studies. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2016 Nov 8:1-8

127 Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol.* 2001 Jun 1;16(6):314-321

128 He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature.* 2010 Mar 25;464(7288):610-4

129 Bandelt HJ, Salas A. Current next generation sequencing technology may not meet forensic standards. *Forensic Sci Int Genet.* 2012 Jan;6(1):143-5

130 Skonieczna K, Malyarchuk B, Jawień A, Marszałek A, Banaszkiewicz Z, Jarmocik P, Borcz M, Bała P, Grzybowski T. Heteroplasmic substitutions in the entire mitochondrial genomes of human colon cells detected by ultra-deep 454 sequencing. *Forensic Sci Int Genet.* 2015 Mar;15:16-20

131 Just RS, Irwin JA, Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci Int Genet.* 2015 Sep;18:131-9